ATTORNEY'S DOCKET NUMBER: 2007651-0001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Jack *et al.*	**Examiner:**	Hutson, Richard G.
Serial No.:	10/089,027	**Art Unit:**	1652
Filing Date:	March 26, 2002	**Confirmation No.:**	9409
Title:	INCORPORATION OF MODIFIED NUCLEOTIDES BY ARCHAEON ANA POLYMERASES AND RELATED METHODS		

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

ELECTRONICALLY FILED VIA EFS WEB

Sir:

APPEAL BRIEF UNDER 37 C.F.R. § 41.37

Appellants hereby appeal to the Board of Patent Appeals and Interferences (the "Board") from the Examiner's final rejection of pending claims 32-42 of the above-referenced application.

A final Office Action was mailed on April 21, 2009. A Notice of Appeal was filed on July 21, 2009. This Appeal Brief is filed on December 23, 2009 with a Petition under 37 C.F.R. § 1.136 for four month extension of time. An electronic payment of the $270.00 fee for filing an appeal brief under 37 C.F.R 41.20(b)(2) and $865 fee for an extension of time is filed concurrently with this submission. Applicant believes that no further petitions and fees are required for this Appeal Brief to be entered. Please consider this a conditional petition for any additional extensions, if needed, and please charge any additional fees or credit any overpayments that may be required to our Deposit Account No. 03-1721 referencing attorney docket number 2007651-0001.

REAL PARTY IN INTEREST

As a result of an assignment by the inventors in the present application, the real party in interest in this application is New England Biolabs, Inc. The assignment to New England Biolabs, Inc. was recorded in the Patent and Trademark Office at Reel 012921, Frame 0103.

RELATED APPEALS AND INTERFERENCES

No other appeals or interferences are known to Appellants, Appellants' legal representative, or Appellants' assignee that will directly affect or be directly affected by the Board's decision in this appeal. Similarly, no such appeals or interferences are known that may have a bearing on the Board's decision in this appeal.

STATUS OF CLAIMS

The application was filed with 31 claims. Various claims were amended and/or cancelled in Amendments filed on June 7, 2002, April 18, 2005 (not entered), August 15, 2005, May 4, 2006 (not entered), May 30, 2006 (not entered), July 3, 2006, and February 24, 2007. Pending claims 2-4, 13-22, and 27-31 were canceled in the amendment filed October 31, 2007, and new claims 32-43 were presented. Claim 32 was amended in an Amendment filed on August 29, 2008 (not entered) and an Amendment filed on January 9, 2009. Claims 32-42 were finally rejected in an Office Action mailed April 21, 2009. Claim 43 is objected to for depending on rejected claims 32 and 33. The rejection of claims 32-42 is hereby appealed. A listing of pending claims 32-43 is provided in the attached **Claims Appendix**.

STATUS OF AMENDMENTS

There are no outstanding amendments to the claims.

SUMMARY OF CLAIMED SUBJECT MATTER

DNA polymerases are enzymes that catalyze polymerization of nucleotides into a DNA strand. The present invention encompasses the finding that a certain class of DNA polymerases has the ability to incorporate a particular modified type of nucleotide, acyclonucleotides, into DNA strands. The present claims therefore recite use of DNA polymerases from that class (defined by the present of a particular amino acid motif whose presence is shown to correlate with the activity) to incorporate acyclonucleotides into a polynucleotide chain.

Independent claim 32 and dependent claims 33-43 specifically recite methods comprising steps of providing a DNA polymerase of the relevant class, contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide, and incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides. The claims require that the utilized DNA polymerase be a member of the relevant class of DNA polymerases by specifying both a level of overall sequence identity to a member of the class and the presence of the correlated motif. Specifically, the claims specify that the DNA polymerase as an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino-acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to three (i.e., 0-3) amino acid substitutions as compared with the SEQ ID NO.

The claimed methods are described, inter alia, in original claims 9, 10; page 19, lines 19-20; page 31, lines 22-28; page 32, lines 1-3; and Table 3 on pages 20-21 of the specification. Support for claim 32 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; page 19, lines 19-20; page 31, lines 22-28; page 32, lines 1-3 and lines 10-16; and Table 3 on pages 20-21. Support for claim 33 is found in the specification as originally filed, inter alia, in original claim 10 and at page 19, lines 18-20. Support for claim 34 found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 35 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 36 is found in the specification as originally filed, inter alia, in

original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 37 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 38 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 39 is found in the specification as originally filed, inter alia, in original claim 19 Support for claim 40 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 41 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 42 is found in the specification as originally filed, inter alia, in original claim 9; page 18, line 30, to page 19, line 2; Table 3 on pages 20-21; and page 32, lines 1-3 and lines 10-16. Support for claim 43 is found in the specification as originally filed, inter alia, in original claims 13 and 18.

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The grounds of rejection to be reviewed on appeal are:

(1) Are claims 32-42 invalid for lack of written description under 35 U.S.C. § 112?

(2) Are claims 32 -42 invalid for lack of enablement under 35 U.S.C. § 112?

GROUPING OF CLAIMS

For reasons discussed below in the Argument section, the claims stand or fall together for purposes of ground of rejection numbered (1) above, as indicated below:

(1) Claims 32 and 39 stand or fall together.

(2) Claim 33 stands or falls alone.

(3) Claim 34 stands or falls alone.

(4) Claim 35 stands or falls alone.

(5) Claim 36 stands or falls alone.

(6) Claim 37 stands or falls alone.

(7) Claim 38 stands or falls alone.

(8) Claim 40 stands or falls alone.

(9) Claim 41 stands or falls alone.

(10) Claim 42 stands or falls alone.

For reasons discussed below in the Argument section, the claims stand or fall together for purposes of ground of rejection numbered (2) above, as indicated below:

(1) Claims 32 and 39 stand or fall together.

(2) Claim 33 stands or falls alone.

(3) Claim 34 stands or falls alone.

(4) Claim 35 stands or falls alone.

(5) Claim 36 stands or falls alone.

(6) Claim 37 stands or falls alone.

(7) Claim 38 stands or falls alone.

(8) Claim 40 stands or falls alone.

(9) Claim 41 stands or falls alone.

(10) Claim 42 stands or falls alone.

ARGUMENT

Ground of Rejection 1:

Claims 32 and 39 are not invalid for lack of written description

Pending claims 32-42 stand rejected for lack of written description. The Examiner states that claims 32-42 contain subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventors, at the time the application was filed, had possession of the claimed invention. This rejection is respectfully traversed. Reconsideration and withdrawal is requested.

The written description requirement serves both to satisfy the inventor's obligation to disclose the technologic knowledge upon which the patent is based, and to demonstrate that the patentee was in possession of the invention that is claimed. *Capon v. Eshhar*, 418 F.3d 1349, 1357 (Fed. Cir. 2005). To satisfy the written description requirement, the applicant does not have to utilize any particular form of disclosure to describe the subject matter claimed, but the description must clearly allow persons of ordinary skill in the art to recognize that he or she invented what is claimed. *Carnegie Mellon Univ. v. Hoffmann La Roche Inc.*, 541 F.3d 1115, 1122 (Fed. Cir. 2008) (quoting *In re Alton*, 76 F.3d 1168 (Fed. Cir. 1996)). In other words, the applicant must 'convey with reasonably clarity to those skilled in the art that, as of the filing date sought, he or she was in possession of the invention,' and demonstrate that by disclosure in the specification of the patent. *Id.* Such disclosure need not recite the claimed invention *in haec verba*, but it must do more than merely disclose that which would render the claimed invention obvious. *Univ. of Rochester v. G.D. Searle & Co.*, 358 F.3d 916, 923 (Fed. Cir. 2004). The descriptive text needed to meet the written description requirement varies with the nature and scope of the invention at issue, and with the scientific and technologic knowledge already in existence. *Capon*, 418 F.3d at 1357.

The present claims recite methods comprising steps of providing a particular type of DNA polymerase, contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide; and incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends

the primer by incorporating the nucleotides. Claim 32 specifies that the DNA polymerase has an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO. The recited 15 amino acid motifs are shown in Table 3 of the specification at pages 20-21.

The Examiner has maintained that the written description requirement is not met for the scope of DNA polymerases encompassed by the claims. Appellants explain below that a structure/function relationship has been established for the DNA polymerases recited in the claimed methods, and that written description for the claims is more than satisfied under *Invitrogen Corp. v. Clontech Labs, Inc.*, 77 USPQ2d 1161 (Fed. Cir. 2005), and under the U.S. Patent and Trademark Office Written Description Training Materials (Revision 1, March 25, 2008).

A Structure/Function Relationship has been established.

The Examiner maintains the rejection for lack of written description on the ground that "applicants have not related the subgenus of structure to the acyclonucleotide incorporation function" (Office Action mailed April 21, 2009, page 4). Appellants respectfully disagree with this assertion. The disclosure of the specification, working examples, and declaratory evidence demonstrates a relationship between the structure recited in the claims and acyclonucleotide incorporation function. The claims require that the DNA polymerase have an amino acid sequence with at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4 (Vent™). The claims also require that the DNA polymerase include a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22, or has up to three amino acid substitutions.

The specification explains that proteins can display sequence similarity over short stretches of primary amino acid sequence (specification, page 14). These patches are thought to occur most often at essential protein interfaces, such as those involved in catalysis, substrate binding, or protein-protein recognition. The degree of sequence similarity, particularly in conserved sequence motifs, is predictive of the degree to which the proteins will behave

similarly in both physical properties and catalytic function (specification, page 14, lines 10-16). The claims include just such a motif, by requiring that the DNA polymerase include a 15 amino acid motif that is identical to one of SEQ ID NOs 5-22, or has up to three amino acid substitutions. The sequences of the 15 amino acid motifs and the DNA polymerase in which each is found are shown in Table 3 of the specification at pages 20-21. Each motif is within a conserved region having a role in substrate binding, known as "motif B" as defined by Delarue et al. (*Protein Eng.* 3:461-467, 1990; see citation to Delarue et al. in the specification at page 21 under Table 3; Delarue et al. was submitted with the Information Disclosure Statement filed on May 9, 2002, and is attached as **Exhibit A**). Delarue et al. does not recognize or discuss acyclonucleotide activity of any DNA polymerases. However, Delarue et al. indeed indicates that motif B is involved in DNA polymerase function. In the Discussion section, Delarue et al. states:

> ***From structure to function*. Considerable biochemical evidence points to the imporance of [motifs A, B and C] in the DNA polymerase activity**. A synthesized *E. coli* pol I oligopeptide corresponding to the N-terminal-most two-thirds of the loop region connecting helices O and P (**motif B**-see Figure 4) **has been shown to bind deoxynucleotide triphosphate substrates of pol I** as well as duplex DNA (Mildvan, 1989)" (Delarue et al., page 465, right col., lines 9-15; emphasis added).

This structure/function relationship between motif B and polymerase activity is further confirmed in declaratory evidence submitted during prosecution of the present application. In the Declaration by Dr. William Jack, filed on May 4, 2006 ("the Jack Declaration"; copy attached as **Exhibit B**), it states that Dr. Jack and colleagues have published articles in peer reviewed journals discussing the physical basis for the preferential incorporation of acyclonucleotides and enhanced incorporation with Vent A488L and 9°N 485L DNA polymerase mutants, citing to Gardner et al. (*J. Biol. Chem.* 279(12):11834-11842, 2004; Gardner et al. was submitted with the Jack Declaration and is attached as **Exhibit B**). Gardner et al. shows an alignment of Family B DNA polymerases in Figure 1. As is clear from the Figure, the "Region III" active site overlaps with the 15 amino acid motif recited in Appellants' claims. As provided in the Jack Declaration, Gardner discusses the physical basis for incorporation of acyclonucleotides at page 11841. This discussion mentions the A288 residue in Vent™, which is in the active site and in the 15 amino acid motif in Appellants' claims. A relationship between

"Region III", containing the 15 amino acid motif, and polymerase function, had been previously noted, e.g., in Hopfner et al. (*Proc. Nat. Acad. Sci. USA* 96:3600-3605, 1999; Hopfner et al. was submitted with the Jack Declaration and is attached as **Exhibit B**). Hopfner et al. reports the crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. Hopfner et al. provide a structure based sequence alignment of archaeal family B polymerases, and show that Region III (which contains the 15 amino acid motif) is in the active site of these enzymes (see Hopfner et al., page 3603, col. 1, Figure 3, and col. 2, third full paragraph). Hopfner et al. discusses the conserved $KX_3NSXYGX_2G$ motif, which is a sub-motif within Appellants' claimed 15 amino acid motif, in the section entitled "Polymerase Active Site", noting that it and a second motif "form the bottom of the nucleotide-binding site" (Hopfner et al., page 3603, right col., lines 31-32). The subgenus of structure (i.e., the 15 amino acid motif) is clearly related to function.

Although there was recognition in the art that conserved motifs found in polymerases are involved in polymerase activity, it is Appellants who recognized and now claim a method of using a specific genus of polymerases which possess acyclonucleotide incorporation function. The set of 15 amino acid motifs specified by SEQ ID NOs 5-22 and recited in the claims are highly related to each other. SEQ ID Nos 6-17 differ from SEQ ID No 5 by three or fewer residues. SEQ ID Nos 18 and 20-22 differ from SEQ ID Nos 5 by six or fewer residues. Motifs of polymerases having sequences sharing less than 30% overall identity with Vent™ (and thus which are outside the scope of the claims) have motifs which differ from SEQ ID No 5 by seven or more residues (see, e.g., SEQ ID Nos 23-30 at page 21, Table 3 of the specification).

A structure/function relationship is not only supported by an understanding of the 15 amino acid motif and its role in enzymatic function. It is also supported by Appellants' working examples. Every DNA polymerase tested that meets the structural requirements of the claims has acyclonucleotide incorporation activity. Indeed, four different DNA polymerases, Vent™, Deep Vent™, *Pfu*, and 9N™, showed the ability to incorporate acyclonucleotides (specification, Example 6). Two variants of these enzymes, Vent™/A488L, and 9N™/A485L, were also shown in incorporate acyclonucleotides (specification, Example 11). By contrast, Thermosequenase, which is a Taq DNA polymerase variant that lacks the 15 amino acid motif required by the claims, showed a much stronger preference for dideoxyoligonucleotides over

acyclonucleotides (specification, Examples 5 and 12). The application therefore establishes the correlation between the sequence motif and the function recited in the claims.

In addition, the Jack Declaration includes an Appendix with data confirming that an archaeon Family B polymerase from *Methanococcus maripaludis*, having 41% sequence identity with Vent DNA polymerase, utilizes acyclonucleotides as a substrate (Jack Declaration Appendix I, attached as **Exhibit B**). Thus, support for a relationship between the DNA polymerases recited in the claims and acyclonucleotide incorporation function has been provided in by information in the specification regarding sequence similarity and function, exemplification of a relationship between the claimed structure and function in the specification, and data and information provided with the Jack Declaration.

The Examiner maintains his rejection without offering any reason *why* the claimed structure/function relationship allegedly has not been established. For example, in the Office Action mailed May 29, 2008, the Examiner said that "[w]hile Applicants comments regarding the homogeneity shared between this group of polymerases continues to be acknowledged, such is acknowledged in light of the degree of the vast majority of DNA polymerases, many of which have a high degree of homogeneity and not all of which share the ability to incorporate acyclonucleotides into a DNA fragment" (Office Action mailed May 29, 2008, page 4). Appellants have related specific structural features (overall sequence identity and the presence of a 15 amino acid motif in the active site of the enzyme) to function (acyclonucleotide incorporation function). The Examiner has provide no *reason* to doubt Appellants correlation. The Examiner is not entitled to substitute his personal skepticism for statements and evidence provided by the Appellants.

Written description support for the claims is met under Invitrogen Corp. v. Clontech Labs, Inc. 77 USPQ2d 1161 (Fed. Cir. 2005).

Relevant legal precedent also confirms that the written description requirement is satisfied for the present claims in view of the present specification. The decision in *Invitrogen Corp. v. Clontech Labs, Inc.* 77 USPQ2d 1161 (Fed. Cir. 2005) requires a finding that the claims

are adequately described. To emphasize this point, Appellants reiterate a close comparison between *Invitrogen* and the present claims here. The claim at issue in *Invitrogen* read:

> 1. An isolated polypeptide having DNA polymerase activity and substantially reduced RNAse H activity, wherein said polypeptide is encoded by a modified reverse transcriptase nucleotide sequence that encodes a modified amino acid sequence resulting in said polypeptide having substantially reduced RNase H activity, and wherein said nucleotide sequence is derived from an organism selected from the groups consisting of a retrovirus, yeast, Neurospora, Drosophila, primates and rodents.

The specification supporting the claim had only a single example of a polymerase having the recited activity. The court found that the claim met the written description requirement because, (1) at the time of the invention, sequences of reverse transcriptase (RT) genes were known; (2) members of the RT gene family shared significant homologies from one species to another; (3) the written description taught that the invention can be applied to RT genes of other retroviruses; and (4) the specification cited references providing the known nucleotide sequences of those genes.

It must be noted that, unlike the claim in *Invitrogen* which recites no structural limitations, the pending claims include explicit recitation of structural features (overall homology and a 15 amino acid motif). The present specification provides six specific examples of DNA polymerases that fall within the claims. As for the other factors from *Invitrogen*, (1) sequences of many DNA polymerases were known when the present application was filed; and (2) members of the DNA polymerase gene family share significant homologies from one species to another. See the present specification, e.g., at page 3, lines 8-21; and page 10, line 12, to page 15, line 34. For (3), the written description of the present case clearly teaches that the invention can be applied to DNA polymerases other than the ones specifically exemplified. See, for example, page 19, lines 15-27, which teaches:

> The similarity of incorporation patterns with these selected enzymes suggests that not only these archaeon DNA polymerases, but a larger family of DNA polymerases could share the ability to incorporate acyclo to a greater extent than dideoxy terminators. Since *Pfu*, Deep Vent® and 9°N™ DNA polymerases have greater than about 70% sequence identity with Vent DNA polymerase, other enzymes with equivalent or greater identity can reasonably be expected to perform as Vent® (exo-) DNA polymerase in this invention. Notably, those

> enzymes for which no significant sequence similarity is found (i.e., Family A DNA polymerases such as Taq) do not perform in similar ways in the current invention. This fact leads us to believe that archaeon enzymes showing intermediate identity, namely those between about 20 and 70% identity are reasonably candidates for this invention.

As to (4), the specification cites references providing the known sequences of such other DNA polymerases (see, for example, page 10, line 22; page 14, line 18; page 14, line 19; page 15, lines 19-24). Moreover, the sequences of other DNA polymerases are known and need not be fully presented in the specification to satisfy the written description requirement. See *Capon*, 418 F.3d at 1358.

Appellants maintain that, with regard to every relevant fact relied upon by the court, the present case has at least as much, or more description than was provided in *Invitrogen*.

The Examiner disputes this point because the claims encompass incorporation of acyclonucleotides into DNA and

> [t]his is not a property of a DNA polymerase that is well known in the art and the applicants have not adequately described this supposedly new function of a specific sub-genus of DNA polymerases. This is in contrast to the claims of Invitrogen in which the homologies of the encompassed DNA polymerases were high and that region responsible for reduced RNAse H activity in each of these DNA polymerases known such that the encompassed DNA polymerase variants known. (Office Action mailed April 21, 2009, page 5).

Appellants explain in detail the relationship between structure provided and acyclonucleotide function above. As discussed, Appellants have demonstrated (through several examples) that DNA polymerases that do have the claimed sequence do have the recited activities, and a DNA polymerase that does not have the claimed sequence does not have the recited activity.

Moreover, the fact that the present claims recite use to perform a newly discovered function (incorporation of acyclonucleotides) does not distinguish the present case from *Invitrogen*, as asserted by the Examiner. The claims in *Invitrogen* also related to DNA polymerases that have a new function (reduced RNAse H activity). The Examiner is correct that the *region* of DNA polymerase sequence that was responsible for RNAse H activity was previously known. As discussed above, the relevant region of DNA polymerases (region III)

involved in the present claims was also known (and known to be important for activity, just not for this activity). The present specification demonstrates that this known region is important for a new activity, much like the specification in *Invitrogen* demonstrated that changes in a known region could reduce activity. Closer factual scenerios in fact would be difficult to find!

Furthermore, Appellants fail to see how acyclonucleotide function of the DNA polymerases renders this case distinguishable from *Invitrogen*. In that case, a single example of an enzyme having a desired function (reduced RNAse H activity) was adequate to support the claims.

Appellants' recognition of a class of polymerases which incorporate acyclonucleotides is new, and Appellants have linked the functional activity with structure and a characterized structural, functional motif (i.e., the 15 amino acid motif). There is no basis for distinguishing the present case from *Invitrogen*. The Examiner suggests that *Invitrogen* is not applicable because "the homologies of the encompassed DNA polymerases were high." Yet the *Invitrogen* claim is completely devoid of structural limitations, and recites polymerases from organisms as diverse as viruses, yeasts, and primates! If unspecified sequences from such varied species have "high" homology in the Examiner's view, Appellants fail to understand how homologies between sequences encompassed by the present claims, which recite concrete structural limitations, are not also "high."

In a further attempt to distinguish *Invitrogen*, the Examiner stated that

> the description held by Invitrogen is specific to the claims of invitrogen [sic], based upon the specification and art as well as a. Actual reduction to practice, b. Disclosure of drawings or structural chemical formulas, c. Sufficient relevant identifying characteristics, such as: Complete structure, ii. Partial structure, iii. Physical and/or chemical properties, iv. Functional characteristics when coupled with a known or disclosed correlation between function and structure, d. Method of making the claimed invention, e. Level of skill and knowledge in the art and f. Predictability in the art. (Office Action mailed April 21, 2009, carryover paragraph from pages 5-6).

Legal decisions would be meaningless as precedent if they could be applied only to a single set of facts. Appellants have provided a close comparison of (i) the facts in the *Invitrogen* case and the (stronger) facts here; and (ii) the claims of the present application and a claim from

Invitrogen for which written description was affirmed. There has been no showing that *Invitrogen's* claimed genus all had "high" homology or "known" function such that the present claims can be distinguished from the case. No other bases for finding *Invitrogen* inapplicable have been offered.

Written description support for the claims is met under the U.S. Patent and Trademark Office Written Description Training Materials

The Examiner compared the present claims to the U.S. Patent and Trademark Office Written Description Training Materials (hereinafter, the "Guidelines") and found lack of description in Appellants' claims compared to claim 2 in Example 11 of the Guidelines because "claim 2 is drawn to a nucleic acid having 85% identity to a specific sequence, a partial structure. This is relative to the instant claims which require even less partial structure of 30% identity." (Office Action mailed April 21, 2009, page 7).

The Examiner has not analyzed Appellants' claims in view of the knowledge of DNA polymerase structure and the requirement of a conserved motif which is associated with enzymatic function. Claim 2 in Example 11 of the Guidelines concerns a claim to nucleic acid encoding hypothetical polypeptide having "activity X". In contrast to the present claims, the hypothetical polypeptide encoded by the nucleic acid does not share significant sequence identity with <u>any</u> known polypeptide or polypeptide family. Also unlike the present claims, the specification for this hypothetical example discloses only a single nucleic acid sequence that encodes a polypeptide having "activity X". Any comparison of the present claims to Example 11 should take these facts into consideration. Another important factor for analysis in Example 11 is the presence of a disclosed or art-recognized correlation between structure and function. Appellants have provided this correlation.

Example 5 of the Guidelines presents a fact pattern much more analogous to Appellants claims, and is a more appropriate basis for comparison. Example 5 concerns a claim to an "isolated protein comprising Protein A," wherein Protein A includes the amino acid sequence of SEQ ID NO:1, has the ability to bind and activate Protein X, and is purified by a recited set of conditions. The sequence of SEQ ID NO:1 in this hypothetical claim has 10 amino acids.

Likewise, Appellants' claims recite DNA polymerases that include a 15 amino acid motif and have a specific binding and activity function, which is the ability to incorporate acyclonucleotides in a polymerase extension reaction. The polymerases are not defined by purification conditions. However, significant structural definition for the polymerases is provided by requiring at least 30% identity to SEQ ID NO:4.

In the hypothetical fact pattern set forth for Example 5, claim 1, the specification fails to disclose the complete structure of Protein A and it fails to disclose any art recognized correlation between the structure of the claimed protein and its function of binding and activating Protein X. Nonetheless, written description is affirmed for the claim because the specification discloses a partial (10 amino acid) sequence of Protein A and because relevant identifying characteristics are provided in the form of its ability to bind and activate Protein X, and purification features.

If anything, the present specification provides more description support for the claims than is provided for claim 1 of Example 5 of the Guidelines. Appellants' specification describes examples of complete structures for polymerases that fall within the claims. Appellants' 15 amino acid motif imposes greater structural definition for a polymerase than the 10 amino acid sequence defining the hypothetical polypeptide of Example 5. Appellants' polymerases possess a binding ability and activity (acyclonucleotide incorporation) which is just as well defined as those of the hypothetical polypeptide of Example 5. Whereas no correlation of protein structure with function is provided in Example 5, Appellants' provide detailed structure/function correlation, as set forth above. In this aspect, Appellants provide more support than the Guidelines require. Another factor favoring support for the hypothetical polypeptide was the specification's disclosure of methods for isolating the polypeptide and a working example showing the polypeptide was successfully isolated. Appellants' have also shown that one of skill in the art can make and use polypeptides as claimed, and that polypeptides have the recited function.

Claim 33 is not invalid for lack of written description

Claim 33 stands rejected for lack of written description. Claim 33 specifies that the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ

ID NO:4. Because this claim requires a higher overall identity to SEQ ID NO:4, the genus of polymerases encompassed by the claim is smaller than that of claim 32. Thus, the level of description required is reduced as compared with claim 32. Appellants' specification demonstrates that multiple polymerases within the genus possess acyclonucleotide function. (Appellants emphasize that polymerases from the broader genus have this function as well; see the Jack Declaration, Appendix I, which shows that a *Methanococcus* DNA polymerase having only 41% sequence identity to Vent™ DNA polymerase incorporates acyclonucleotides more efficiently than dideoxynucleotides.) Even if claim 32 were not fully supported by the specification (which Appellants do not concede), claim 33 would be.

Claim 34 is not invalid for lack of written description

Claim 34 stands rejected for lack of written description. Claim 34 specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-22. Given the further limitation on the sequence of the motif (i.e., such that the motif does not include amino acid substitutions), the genus of polymerases encompassed by the claim is smaller than that of claim 33. The level of description required for this claim is reduced as compared with claim 32. Even if claim 32 were not fully supported by the specification, claim 34 would be.

Claim 35 is not invalid for lack of written description

Claim 35 stands rejected for lack of written description. Claim 35 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 15-17, except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO. Because it covers fewer motifs, this claim refers to a genus of polymerases that is smaller than that encompassed by claim 32. The level of description required to support this claim is less than required for claim 32.

Claim 36 is not invalid for lack of written description

Claim 36 stands rejected for lack of written description. Claim 36 specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-17. The genus of polymerases

encompassed by this claim is even smaller than that of claim 32 and requires less description to be adequately supported.

Claim 37 is not invalid for lack of written description

Claim 37 stands rejected for lack of written description. Claim 37 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it may contain up to three amino acid substitutions. Again, the genus of polymerases encompassed by this claim is even smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully supported by the specification.

Claim 38 is not invalid for lack of written description

Claim 38 stands rejected for lack of written description. Claim 38 specifies that the amino acid motif is identical to one of SEQ ID NOs 5-8. The genus of polymerases encompassed by this claim is smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully supported by the specification.

Claim 40 is not invalid for lack of written description

Claim 40 stands rejected for lack of written description. Claim 40 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

Claim 41 is not invalid for lack of written description

Claim 41 stands rejected for lack of written description. Claim 41 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

Claim 42 is not invalid for lack of written description

Claim 42 stands rejected for lack of written description. Claim 42 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID Nos 5-8. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully supported by the specification.

In conclusion, the provided teachings in the specification, examples, sequences, declaratory evidence, and data are more than sufficient to describe function and support description of the claims. The Examiner has not established otherwise. For reasons set forth above, withdrawal of the rejection of claims 32 and 39 as allegedly lacking written description is respectfully requested.

Ground of Rejection 2:

Claims 32 and 39 are not invalid for lack of enablement

Pending claims 32 and 39 stand rejected for lack of enablement. The Examiner states that the specification, while being enabling for a method comprising providing a DNA polymerase selected from the group consisting of Vent™, Deep Vent™, *Pfu*, and 9°TM or the specifically disclosed variants of claim 43, "does not reasonably provide enablement for any method comprising providing a DNA Polymerase having an amino acid sequence that shows a mere 30% overall identity with that of SEQ ID NO:4 and further includes a 15 amino-acid motif that is identical to SEQ ID NO:5 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO…" (Office Action mailed April 21, 2009, pages 8-9). The Examiner stated that "determination of those DNA polymerases having the desired biological characteristics is unpredictable and the experimentation left to those skilled in the art is unnecessarily, and improperly, extensive and undue" (Final Office Action mailed April 21, 2009, pages 11). Appellants have previously reviewed the factors set forth in *In re Wands* (858 F.2d 731, 8 USPQ2nd 1400, Fed. Cir. 1988) with respect to the present claims and review them here, in response to the Examiner's assertion that Appellants' burden has not been met.

First, Appellants address the Examiner's comments regarding *Wands* factor (2), Amount of Direction or Guidance. In the Final Office Action mailed April 21, 2009, the Examiner maintained that guidance was lacking as to DNA polymerases which have the ability to incorporate acyclonucleotides into a DNA template, and requested clarification as to how the 15 amino acid motif correlates with acyclonucleotide function (Office Action mailed April 21, 2009, page 12). As explained above in the arguments for written description, the 15 amino acid motif is a highly conserved motif in the active site of family B DNA polymerases which plays a role in substrate binding. The Examiner disputes a structure/function correlation because "applicants have not disclosed such a single motif but rather continue to refer to any of a number of motifs or variants thereof"(Office Action mailed April 21, 2009, page 12). Some variability within the genus of motifs is permitted, given that variable polymerases share acyclonucleotide incorporation function. For example, both 9°N polymerase and Vent™ incorporate

acyclonucleotides, although their 15 amino acid motifs differ by three amino acids (compare SEQ ID NO 5 and SEQ ID NO 7 at page 20, Table 3 of the specification). A *Methanococcus maripaludis* DNA polymerase having a more divergent sequence also possesses acyclonucleotide incorporation activity. The claims do require a degree of conservation of sequence, which is clearly expressed in the claims. The fact that variable polymerases share a specific function does not render them "unpredictable."

As to the (1) Quantity of Experimentation Necessary, and (3) Presence or Absence of Working Examples, Appellants reiterate that one or ordinary skill could make and test all polypeptides within the scope of the claims to determine their ability to extend a DNA primer or incorporate acyclonucleotides (including to determine their ability to preferentially select acyclonucleotides). Appellants' working examples include demonstration of activity of multiple species of DNA polymerases set forth in the specification and in declaratory evidence discussed herein.

As to (5) State of the Prior Art, and (7) Predictability of the Art, Appellants note, and the Examiner has acknowledged, that the prior art with regard to DNA polymerases and their classification is extensive. However, Appellants disagree with the Examiner's assertion that "determination of those DNA polymerases having the desired biological characteristics is unpredictable and the experimentation left to those skilled in the art is unnecessarily, and improperly, extensive and undue" (Office Action mailed April 21, 2009, page 11). Appellants have identified polymerases which have acyclonucleotide incorporation function by virtue of structural and physical characteristics distinctive of well-characterized DNA polymerases. These characteristics include overall sequence identity to a polymerase, and the presence of a conserved motif. Appellants have shown that all members tested within the genus of polymerases have the recited activity. The Examiner's only discernible reason for declaring these features unpredictable is the breadth of the genus of polymerases. This improperly disregards Appellants' demonstration of activity for multiple species. It also disregards the fit of Appellants' observed activity with well characterized classification schemes for DNA polymerases (among which one finds substantial variability despite conserved nucleotide polymerase activity).

As to the (4) Nature of the Invention and (8) Breadth of the Claims, Appellants reiterate that DNA extension reactions are well within the skill of those of ordinary skill. As part of their invention, Appellants have described a class of DNA polymerases that can incorporate acyclonucleotides, and have shown function for six different species within the class. Given the demonstrated correlation of structure with function and other reasons provided above, Appellants disagree with the Examiner's assertions that the scope of the claims is not enabled.

As to (6) Relative Skill of those in the Art, Appellants submit, and the Examiner has agreed, that the relative skill of those in the art is very high.

Claim 33 is not invalid for lack of enablement

Claim 33 stands rejected for lack of enablement. Claim 33 depends from claim 32 and specifies that the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ ID NO:4. Because this claim requires a higher overall identity to SEQ ID NO:4, the breadth of the claim is smaller than that of claim 32. The scope of enablement provided by the disclosure is more than sufficient to support the scope of this claim, not least because multiple polymerases that fall within the claimed genus are exemplified.

Claim 34 is not invalid for lack of enablement

Claim 34 stands rejected for lack of enablement. Claim 34 depends from claim 32 or 33 and specifies that the 15 amino acid motif is identical to one of SEQ ID Nos 5-22. This further limitation on the sequence of the motif (i.e., such that the motif does not include amino acid substitutions) provides a claim of smaller breadth than claim 32 and which is more than supported by the disclosure.

Claim 35 is not invalid for lack of enablement

Claim 35 stands rejected for lack of enablement. Claim 35 depends from claim 32 or 33 and specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 15-17, except that it

contains up to 3 amino acid substitutions as compared with the SEQ ID NO. This claim covers fewer motifs than claim 32 and is enabled for its full scope.

Claim 36 is not invalid for lack of enablement

Claim 36 stands rejected for lack of enablement. Claim 36 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-17. Again, the genus of polymerases encompassed by this claim is even smaller than that of claim 32 and is enabled by the disclosure provided.

Claim 37 is not invalid for lack of enablement

Claim 37 stands rejected for lack of enablement. Claim 37 specifies that the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it may contain up to three amino acid substitutions. The genus of polymerases encompassed by this claim is even smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully enabled by the specification.

Claim 38 is not invalid for lack of enablement

Claim 38 stands rejected for lack of enablement. Claim 38 specifies that the amino acid motif is identical to one of SEQ ID NOs 5-8. The genus of polymerases encompassed by this claim is smaller than that of claim 32 due to further limitation of the 15 amino acid motif and is fully enabled by the specification.

Claim 40 is not invalid for lack of enablement

Claim 40 stands rejected for lack of enablement. Claim 40 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully enabled by the specification.

Claim 41 is not invalid for lack of enablement

Claim 41 stands rejected for lack of enablement. Claim 41 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is fully enabled by the specification.

Claim 42 is not invalid for lack of enablement

Claim 42 stands rejected for lack of enablement. Claim 42 specifies that the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID Nos 5-8. The genus of polymerases encompassed by this claim is also smaller than that of claim 32 and is enabled for its full scope.

In light of the above, Appellants submit that claims 32 -42 satisfy the enablement requirement. Allowance of the claims is requested.

Respectfully submitted,

Date: December 23, 2009

/Margo H. Furman/
Margo H. Furman, PhD, JD
Registration Number 59,812
Tel. No. (617) 248-4073

PATENT GROUP
CHOATE, HALL & STEWART LLP
Two International Place
Boston, MA 02110
Tel: (617) 248-5000
Fax: (617) 502-5002
patentdocket@choate.com

CLAIMS APPENDIX

1-31. (Canceled)

32. (Previously presented) A method comprising steps of:

providing a DNA polymerase having an amino acid sequence that shows at least 30% overall identity with that of the polypeptide encoded by SEQ ID NO:4, and further includes a 15 amino-acid motif that is identical to one of SEQ ID NOs 5-22 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO;

contacting the DNA polymerase with a template, a primer that binds to the template, and a collection of nucleotides including at least one acyclonucleotide; and

incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides.

33. (Previously presented) The method of claim 32, wherein the DNA polymerase has an amino acid sequence that shows at least 70% overall identity with that of SEQ ID NO:4.

34. (Previously presented) The method of claim 32 or claim 33, wherein the 15 amino-acid motif is identical to one of SEQ ID NOs 5-22.

35. (Previously presented) The method of claim 32 or claim 33, wherein the 15 amino-acid motif is identical to one of SEQ ID NOs 5-17 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO.

36. (Previously presented) The method of claim 35, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-17.

37. (Previously presented) The method of claim 32 or 33, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-8 except that it contains up to 3 amino acid substitutions as compared with the SEQ ID NO.

38. (Previously presented) The method of claim 37, wherein the 15 amino acid motif is identical to one of SEQ ID NOs 5-8.

39. (Previously presented) The method of claim 32 or 33, wherein the step of incubating comprises incubating the DNA polymerase with the template and the nucleotides so that the DNA polymerase extends the primer by incorporating the nucleotides, and preferentially incorporates acyclonucleotides.

40. (Previously presented) The method of claim 32 or 33, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-22.

41. (Previously presented) The method of claim 35, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-17.

42. (Previously presented) The method of claim 37, wherein the 15 amino acid motif has up to one amino acid substitution as compared with one of SEQ ID NOs 5-8.

43. (Previously presented) The method of claim 32 or 33 wherein the DNA polymerase is Vent™, Deep Vent™, 9°N, *Pfu*, Vent™/488L, or 9°N/485L.

EVIDENCE APPENDIX

Appellants had provided the following evidence during prosecution of the instant application:

Exhibit A: Delarue et al., *Protein Eng.* 3:461-467, 1990. This reference was cited in the Information Disclosure Statement and Form PTO-1449 filed on May 9, 2002, and was entered into the record on May 13, 2002. The Form PTO-1449 was initialed by the Examiner on September 29, 2004, confirming that the reference was entered into the record.

Delarue et al. is attached hereto at pages 33-39.

Exhibit B: Declaration of William Jack, accompanying references and Appendix I. The Declaration was submitted with four references, listed below, and Appendix I along with a response to Office Action filed May 4, 2006, and was entered into the record in PAIR on May 9, 2006 as the entry designated "Rule 130, 131 or 132 Affidavits." Entrance into the record was confirmed by the Examiner's reference to this Declaration on page 3 of the Advisory Action mailed on July 5, 2006.

The Declaration of William Jack is attached hereto at pages 40-47.

Rodriguez et al., *J. Mol. Biol.* 299:447-462, 2000, is attached hereto at pages 48-63.

Gardner et al., *J. Biol. Chem.* 279(12): 11834-11842, 2004, is attached hereto at pages 64-72.

Hashimoto et al., *J. Mol. Biol.* 306:469-477, 2001, is attached hereto at pages 73-81.

Zhao et al., *Structure* 7(10):1189-1199, 1999, is attached hereto at pages 82-92.

Hopfner et al., *Proc. Nat. Acad. Sci. USA* 96:3600-3605, 1999, is attached hereto at pages 93-98.

Appendix I is attached hereto at pages 99-101.

RELATED PROCEEDINGS APPENDIX

Not applicable.

EXHIBIT A

Protein Engineering vol.3 no.6 pp.495–497, 1990

An attempt to unify the structure of polymerases

Marc Delarue, Olivier Poch[1], Noel Tordo[2], Dino Moras and Patrick Argos[3]

Laboratoire de Cristallographie and [1]Biochimie, IBMC du CNRS, 15 rue René Descartes, 67084 Strasbourg, [2]Service Rage Recherche, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France and [3]European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, 6900 Heidelberg, FRG

With the great availability of sequences from RNA- and DNA-dependent RNA and DNA polymerases, it has become possible to delineate a few highly conserved regions for various polymerase types. In this work a DNA polymerase sequence from bacteriophage SPO2 was found to be homologous to the polymerase domain of the Klenow fragment of polymerase I from *Escherichia coli*, which is known to be closely related to those from *Staphylococcus gonococcus*, *Thermus aquaticus* and bacteriophages T7 and T5. The alignment of the SPO2 polymerase with the other five sequences considerably narrowed the conserved motifs in these proteins. Three of the motifs matched reasonably all the conserved motifs of another DNA polymerase type, characterized by human polymerase α. It is also possible to find these three motifs in mitochondrial DNA-dependent RNA polymerases and two of them in DNA polymerase β and DNA terminal transferases. These latter two motifs also matched two of the four motifs recently identified in 14 RNA-dependent polymerases. From the known tertiary architecture of the Klenow fragment of *E.coli* pol I, a spatial arrangement can be implied for these motifs. In addition, numerous biochemical experiments suggesting a role for the motifs in a common function (dNTP binding) also support these inferences. This speculative hypothesis, attempting to unify polymerase structure at least locally, if not globally, under the pol I fold, should provide a useful model to direct mutagenesis experiments to probe template and substrate specificity in polymerases.

Key words: catalytic domain/DNA polymerases/RNA polymerases/sequence alignment

Introduction

The number of available protein sequences is growing rapidly due to the facility of nucleotide sequencing techniques. One protein class often studied and sequenced is the polymerase family which is central to the duplication and expression of genes. Polymerases can use RNA or DNA as a template (RNA- or DNA-dependent); the product can also be RNA or DNA. Polymerases are found both in eukaryotes and prokaryotes, though sequencing efforts have often concentrated on those from viruses. One way to use the information contained in all these sequences is to try to align them and thereby allocate them amongst related families and subfamilies. This has been achieved for DNA-dependent DNA polymerases, where three main subfamilies have been identified. One of them contains the Klenow fragment of *Escherichia coli* polymerase I, whose three-dimensional structure is known (Ollis *et al.*, 1985a), and polymerases from phages T7 (Ollis *et al.*, 1985b; Argos *et al.*, 1986) and T5 (Leavitt and Ito, 1989), and from *Herpesvirus saimiriae* (Lawyer *et al.*, 1989) and *Saccharomyces pneumoniae* (Lopez *et al.*, 1990). This family will be referred to as the pol I family. For another set of DNA-dependent DNA polymerases (Wang *et al.*, 1989), as those homologous to the human polymerase α (hereafter referred to as pol α), more than 10 sequences from various species are known. A third subfamily of DNA-dependent DNA polymerase, hereafter called the pol β type, has only two members: DNA polymerase β (Matsukage *et al.*, 1987) and terminal transferase (Peterson *et al.*, 1985; Kotani, 1986; Zmudzka *et al.*, 1986). Until now, just one DNA-dependent DNA polymerase sequence, from the SPO2 bacteriophage (Besley and Rutberg, 1984; Jung *et al.*, 1987), resisted alignment with any of the three aforementioned types.

Clearly, the aim of these alignments, apart from evolutionary implications, is the identification of the regions essential for polymerase function, since these sequence segments should appear as the most conserved. Generally, a great number of aligned sequences will ensure sufficient variability to identify the functionally required regions. For the pol I type, the five previously aligned (Ollis *et al.*, 1985b; Argos *et al.*, 1986; Leavitt and Ito, 1988; Lopez *et al.*, 1990) sequences are sufficiently close so as not to allow confident delineation of the absolutely required motifs.

In the present work, it has been found that the polymerase from bacteriophage SPO2 can be aligned, using a sensitive method, with the polymerase portion of the Klenow fragment in the C-terminal part of the protein [the N-terminal domain has a 3′–5′ exonuclease function (Freemont *et al.*, 1988)]. The total alignment of the C-terminal part of the six proteins of the pol I type is presented. The relatedness of SPO2 polymerase with those from phages T7 and T5, *Lymphocryptovirus*, *Thermus* and *E.coli* is sufficiently distant that highly conserved regions can now be reduced to five in number. Interestingly, three of the five regions match reasonably with the three most conserved motifs of DNA-dependent DNA pol αs, suggesting that the two polymerase types may share a common tertiary fold, or at least contain similar local tertiary architecture required for similar functions. These motifs are likely to represent modules required for the polymerase structure and activity.

Searches were then performed to detect such sequence patterns in other polymerase families. All three motifs could be found in DNA-dependent RNA polymerases that consist of only one subunit (see Mesters *et al.*, 1987); two motifs were found in pol βs, in the same linear arrangement and maintaining the strictly conserved residues. In addition to this, an examination of several aligned RNA-dependent RNA polymerases as well as reverse transcriptases, for which four highly conserved motifs have been highlighted (Kamer and Argos, 1984; Poch *et al.*, 1989), allowed the suggestion that two of these motifs could match the two motifs shared by DNA pol Is, pol αs and pol βs. These sequence similarities are further supported by a statistically significant alignment between the entire polymerase domain of two members of these different families, namely a DNA-dependent DNA

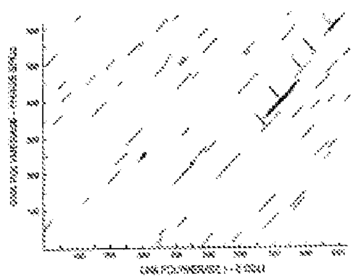


Fig. 1. Homology search matrix (Argos, 1987) between the Klenow fragment of DNA poly I from E.coli and DNA polymerase from bacteriophage SPO2. The search window lengths ranged from 5 to 25 in steps of 2; the search peaks are plotted over the entire window length, with the largest value (excluding other overlap) from the ... windows. The peak values (S) are scored as a number of standard deviations (σ) above the random mean for each window length. Thin lines indicate 4.5σ < S < 5.5σ and thick lines refer to 5.5σ < S < 6.0σ. The path used as a basis to effect the alignment of Figure 2 is indicated by arrows. The remaining alignments shown in Figure 2 are founded on simultaneous inspection of all six sequences.

polymerase and an RNA-dependent DNA polymerase. The alignment of the C-terminal most motif, containing two acidic groups, is in agreement with those suggested by Argos (1988) amongst RNA-dependent RNA polymerases, reverse transcriptases, and DNA-dependent DNA pol αs.

An examination of these conserved regions relative to the known Klenow fragment tertiary architecture positions them in areas likely to interact with nucleotides. In fact, many biochemical data (Joyce and Steitz, 1987; Leavitt et al., 1987b) point to the importance of these motifs in the catalytic process, in the different families of polymerases. The design of site-directed mutagenesis experiments should benefit from the model suggested here, implying possible common tertiary segments for various polymerases, if not a common topology.

Results

Pairwise sequence comparisons were done by a search procedure based on residue characteristics (Argos, 1987). The resulting alignments of the C-terminal polymerase portion of the Klenow fragment of DNA-dependent DNA polymerase I from E.coli and polymerases from bacteriophages T7, T5, S.pneumoniae and T.aquaticus are similar to the ones given by Argos et al. (1986), Leavitt and Ito (1989), Lopez et al. (1989) and Lawyer et al. (1989) respectively. The DNA polymerase sequence from bacteriophage SPO2 was also compared to the E.coli, S.pneumoniae, T5 and T7 polymerases. The search matrix and

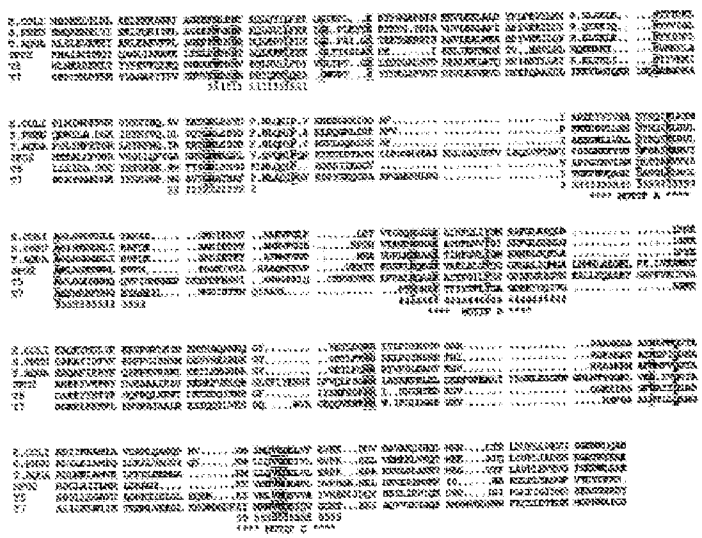


Fig. 2. Multiple sequence alignment of the C-terminal regions of E.coli DNA polymerase I, S.pneumoniae, T.aquaticus, bacteriophages SPO2, T5 and T7 polymerases. The major conserved motifs are numbered 1–5. The number of the first amino acid of the displayed sequence is: 556, 511, 455, 215, 352 and 339 for E.coli, S.pneumoniae, T.aquaticus, SPO2, T5 and T7 respectively.

alignment pattern when the SPO2 and *E.coli* polymerases are shown in Figure 1; it is clear that strong regions of homology exist at the 4.5 SD or higher level. No such strong relationships could be found between SPO2 and *S.pneumoniae*, *T.aquaticus*, T7 or T5 polymerases. Figure 2 shows the alignment of the C-terminal part of these six polymerase sequences; this multiple alignment was obtained by manual adjustment of the different but closest pairwise alignments. A conservation profile resulting from this alignment was also calculated (data not shown); this profile is based on a five-residue window and the score is simply the normalized sum of the matrix elements corresponding to the mutations observed in all the different pairwise alignments. The

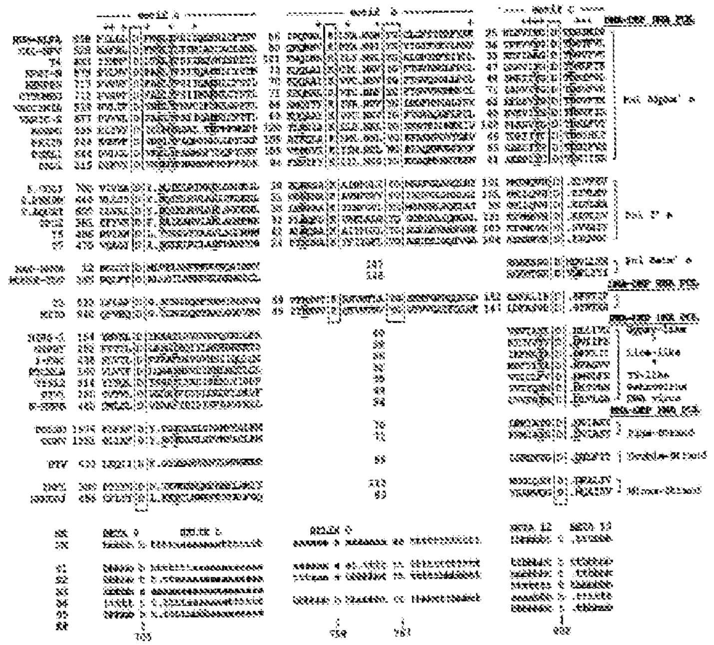


Fig. 3. Multiple alignment of various polymerase sequences for motifs A, B and C. [illegible]

USSN 10/089,027

Atty. Docket No. 2007651-0001

4567604v1

matrix is a rescaled Dayhoff matrix as used by Gribskov and Burgess (1986). A threshold of 60% conservation leaves five regions of high homology, which are indicated in Figure 2. These regions are characterized by the conserved sequences (F)(S)**(S)**(Q)(L)***L for the first region [* refers to a given alignment position occupied by any amino acid, (X) indicates an amino acid almost universally conserved by the six sequences, and a single letter refers to exact conservation in all the sequences], (T)GR for region 2, D*(S)**E for region 3, K*****Y(G) for region 4 and VHD(E) for region 5.

A multiple alignment was also built for 12 sequences of DNA-dependent DNA polymerases α from man, yeast and several viruses. These sequences are essentially the ones presented by Wang et al. (1989) together with two recent additions to the NBRF sequence data bank (one polymerase from another adenovirus, Wood12, and one from a nuclear polyhedrosis virus, Npolhepoval). The resulting conservation profile using a window of five residues left three regions above a 60% conservation threshold. They are characterized by D**(S)(L)(Y)(P)(S), K***(S)(S)*(Y)G and (Y)(G)(D)TDS, as previously noted by other authors (Bernad et al., 1987). A comparison of the most conserved sequences for the pol I and pol α types suggests an overlap of these three regions of pol α with regions 3, 4 and 5 of the pol I type; these regions will hereafter be called motifs A, B and C (Figure 3). These three motifs are centered on invariant residues: motif A contains a strictly conserved aspartate at the junction of a beta strand and an alpha helix, motif B contains an alpha helix with positive charges and motif C has a doublet of negative charges located in a beta–turn–beta secondary structure, as implied from the Klenow structure (Figure 4). Great variability in the lengths separating the different conserved regions was observed for the pol α: 50–120 residues between motifs A and B (average 70) and 25–130 residues between motifs B and C (average 60). In the pol Is, these values compare respectively with 30 and 100 residues, on average.

A rat DNA-dependent DNA polymerase β (Matsukage et al., 1987) can be easily aligned with a mouse terminal deoxynucleotidyltransferase (Koiwai et al., 1986). Though many residues are identically conserved amongst the two sequences, an N-terminal and a C-terminal fragment may be related to motifs A and C of DNA polymerases (Figure 3).

For DNA-dependent RNA polymerases, the three motifs could only be found in the proteins made of only one subunit. Four different sequences of this type are known, the two most distantly related being from phage T3 and yeast mitochondria (Masters et al., 1987). The distances between motifs compare well with the ones of the pol I type. The sequence fragments corresponding to the motifs are presented in Figure 3, together with those found in DNA pol αs, pol Is and pol βs.

Secondary structure predictions averaged over each set of sequences (DNA pol. αs, pol Is and pol βs and DNA-dependent RNA pols) were made according to the procedures of Argos (1985) and are given in Figure 3. Except for pol βs, for which only two sequences are known, they often show agreement amongst themselves as well as with the pol I structure, especially for motifs A and C.

Recently, Poch et al. (1989) identified four motifs shared by all RNA-dependent RNA, and DNA polymerases (84 different sequences). Two of these motifs were found to be similar to two of those contained in the DNA-dependent polymerases; namely, motifs A and C. These two motifs maintain the strictly conserved residues in the same environment. The mean secondary structure predictions agree with the observed structure in the corresponding

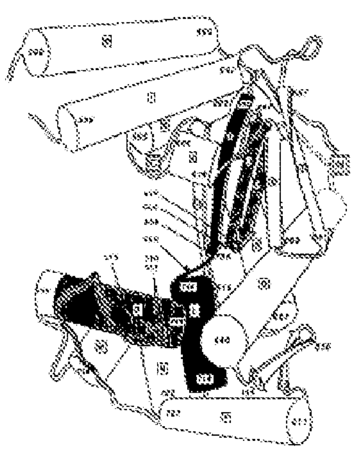


Fig. 4. An illustration of the enzyme folding pattern of the polymerase domain from the E.coli Klenow fragment. The sketch is similar to that given by Ollis et al. (1985b). The three differently shaded regions, in moving from the N-terminus to the C-terminus, correspond respectively to motifs A (black region), B (vertical stripes) and C (horizontal stripes), whose sequences are given in Figure 3. Sequence positions relative to the secondary structures are indicated. β-strands are numbered sequentially and α-helices are designated by letters according to the alphabet. The approximate position of the conserved amino acids is indicated in each of the motifs.

regions of the Klenow fragment (Poch et al., 1989). Sequence fragments from 12 representative RNA-dependent polymerases (i.e. from each of the subfamilies: reverse transcriptases, plus-, minus- and double-strand RNA polymerases) are shown aligned to those matched in the DNA pols (Figure 3).

Several viral RNA-dependent RNA polymerase and reverse transcriptase sequences were then compared to those from pol Is and pol αs using the same sequence comparison technique as above (Argos, 1987). Among the different pairwise comparisons, extensive sequence similarities were found between the viral hepatitis B RNA-dependent DNA polymerase from Woodchuck (Galibert et al., 1982) and Herpes simplex virus DNA polymerase, a member of the DNA pol α family (Gibbs et al., 1985). The matrix is shown in Figure 5 and the resultant alignment in Figure 6. It is noteworthy that 31% of the aligned residues are identical. In this alignment, two conserved regions can be delineated that correspond to motifs A and C. The homology between these two sequences provides a possible link between the DNA- and RNA-dependent polymerases.

Discussion

DNA-dependent DNA polymerases

Sequence similarities. In this work, a new member was added to the pol I type of DNA-dependent DNA polymerases. An analysis of the multiple alignments for the pol I type as well as for the pol α type pointed to three conserved motifs held in

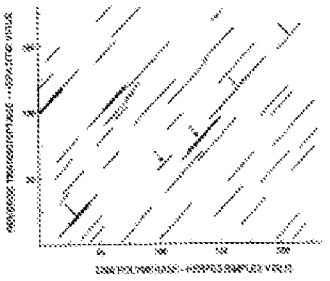


Fig. 8. Homology search (Argos, 1987) between Wondrack (Segment 8 RNA-dependent DNA polymerase and Herpes simplex DNA-dependent DNA polymerase. The search window length ranged from 9 to 35 in steps of 2; the search peaks are plotted over the entire window length, with the longest value corresponding when overlap occurs from the subsequences. The peak values (P) are scored as a number of standard deviations (σ) above the random mean for each window length. The three intervals $3.0\sigma < S < 3.8\sigma$; thick lines, $4.0\sigma < S < 4.5\sigma$; circles $4.6\sigma < S < 4.9\sigma$; and jagged lines $5.0\sigma < S < 5.6\sigma$. The peak used as a basis to effect the alignment shown in Figure 6 is indicated by arrows.

sequence. The distances between these motifs are quite variable. However, within each polymerase type, there is already great variation in the distances between conserved regions. Moreover, it would seem unlikely that those of five most conserved sequence regions in the polymerase domain of pol I (Figure 2) could be aligned with the three most conserved regions of DNA pol α purely by chance. A different piece of evidence pointing to a possible link between the pol I and the pol α types comes from sequence similarities within the exonuclease domain of the T5 polymerase and pol α (Leavitt and Ito, 1989). It is then reasonable to expect their polymerase domains to be related also. In fact, this observation has been recently extended to all exonuclease domains of pol α and pol I's, including SPO2 (Bernad *et al.*, 1989). In addition, the secondary structure predictions, especially for motifs A and C, point to superimposable structures for pol α's that coincide with those observed in the pol I Klenow structure.

From the sequences to structure. The three motifs A, B and C are clustered and consecutive in the C-terminal fraction of the Klenow fragment and correspond to structural features likely to interact with the DNA (Ollis *et al.*, 1985b). Figure 4 highlights the portions of the *E. coli* pol I tertiary structure corresponding to the three shared motifs. Motif A, characterized by a conserved Asp and Tyr, corresponds to strand 9 and helix L. Motif B, with conserved Lys, Tyr and Gly encompasses most of helix O and the following long loop region. Motif C deliminates strands 12 and 13 with the evenly conserved negative charges contained in the connecting loop of the β-hairpin. Since the N-terminal Asp of the doublet is the only one universally conserved, it is presumed to be catalytically more important. It is clear that the structural segments A and B are likely to come into contact with DNA and that the strand 12 – 13 hairpin loop could place at least one catalytic residue in the polymerase active site. Model building of DNA into the Klenow fragment structure also suggests the importance of these regions (Ollis *et al.*, 1985a,b; Warwicker *et al.*, 1985). Furthermore, the C_α atom of motif C Asp882 is within 3.5 Å of that from Asp705, conserved in motif A (C_α atom coordinates given in Brookhaven database (Bernstein *et al.*, 1977, file 1DPE). Their spatial proximity would allow both to participate in catalysis. Region 2 of pol II (sequence TGR between strands 7 and 8—see Figures 3 and 4) is also located in the vicinity of this area but no sequence homology with the pol α type could be found for this region. Region 1 falls in an undefined part of the electron density loop (see Ollis *et al.*, 1985a; Figure 4).

From structure to function. Considerable biochemical evidence points to the importance of these three motifs in the DNA polymerase activity. A synthesized *E. coli* pol I oligopeptide corresponding to the N-terminal most two-thirds of the loop region connecting helices O and P (motif B—see Figure 4) has been shown to bind deoxynucleoside triphosphate substrates of pol I as well as oxoden DNA (Modak, 1989). Furthermore, photo-affinity labeling with 8-azido-dATP identifies Tyr766 as a residue in the active site (Joyce and Steitz, 1987) while Lys758, also part of this motif, succumbs to chemical labeling. This was using pyridoxal phosphate (Basu and Modak, 1987). A further synthesized peptide (Bhattacharyya *et al.*, 1990; Modavan, 1989) corresponding to helix Q and strands 12 and 13 (motif C) was not found to bind the pol I substrate, although it apparently retains its proper folding; however, this result does not exclude that this peptide, although unable to bind dNTP on its own, can correspond with other regions of the native protein and be part of the binding site. In fact, His881 of pol I, which is in the loop joining strands 12 and 13 and sequentially adjacent to the conserved Asp882 of motif C, has been shown to be involved in the binding of [^{32}P]ATTP (Pandey *et al.*, 1987). In addition, the beginning of helix Q, which is close in space to the end of helix O (motif B) and motifs A and C, can also be labeled by phenylglyoxal (Modavan *et al.*, 1988).

Similarly, for the pol α type, genetic studies of the Herpes Simplex virus polymerase (Larder *et al.*, 1987a) revealed that four out of the six identified drug-resistant polymerase mutants cluster in motifs A and B. The drug used was a dNTP analog. Other mutants sensitized to drug and substrate recognition were also mapped by Gibbs *et al.* (1988) in motifs A and B. This is in agreement with a central role of these motifs in dNTP binding for the pol α type, as is the case of the pol I type.

Finally, for the first DNA polymerase type (pol β), a chemical affinity labeled [^{32}P]8-azido-dATP) peptide to the terminal transferase allowed the mapping of the dNTP binding site; this peptide contained the sequence GHEWD that is part of motif C (Evans *et al.*, 1989) and that resembles both YGDTD and WSHDE. This supports the alignment given in Figure 3, even though motif B could not be found with certainty in this pol β type of DNA polymerase. Furthermore, the number of residues between motifs A and C in the transferase (~ 150) matches well the corresponding number in the Klenow pol (a averaging about 120 residues).

Structural implications of the distance variability between motifs. Figure 5 shows the number of residues contained between each of the motifs for all the DNA pol α sequences and for the Klenow polymerase I domains. Between motifs A and B a comparable number of residues is found for both pol I and α types, while between B and C the pol α sequences generally contain considerably fewer amino acids than their putative Klenow counterparts. This region of the structure encompasses helix P, strands 10 and 11, and helix Q. Apart from the beginning of helix Q, these regions are unlikely to be in contact with the DNA, according to model building studies (Warwicker *et al.*, 1985; Figure 4), suggesting that this region could be considerably

HepWoo	[illegible]
Herpes	[illegible]
HepWoo	[illegible]
Herpes	[illegible]
HepWoo	[illegible]
Herpes	[illegible]

Fig. 5. Alignment of hepatitis B reverse transcriptase (RNA-dependent DNA polymerase) from Woodchuck (HepWoo) with the DNA-dependent DNA polymerase from Herpes simplex virus (Herpes). (*), residue identity; (:), an insertion in the Herpes sequence, namely, in single letter code, [illegible].

deleted without apparent catalytic harm. The shortest segment between motifs B and C is 25 amino acids in the human pol α sequence; a peptide of this size could easily connect Ile779 at the C-terminus of the Klenow motif B to Arg835 at the N-terminus of motif C; i.e. connect the loop following helix O with strands 12 and 13 of motif C. In fact, given the 49 Å distance between the C_α atoms of Ile779 and Arg835 in the *E.coli* pol I tertiary structure (Ollis et al., 1985a), 17 residues in helical conformation (mean C_α distance of 1.5 Å) and eight in a coil structure (mean C_α distance of 2.9 Å) could span the required 49 Å. This 'helical model' would allow the maintenance of the N-terminal region of helix Q, whose sequence is also reasonably conserved in the six type I DNA polymerase sequences (see Figure 2).

The distance between motifs A and B can also vary from 58 to 120 residues in the pol α family; this region corresponds to helices M and N of the Klenow structure. In the pol I family, this distance ranges from 39 to 48 residues. An inspection of Figure 4 suggests the possibility of insertions in this region.

DNA-dependent RNA polymerases

In this family, three conserved regions that matched the three motifs A, B and C could be found. Invariant residues are maintained and the residue lengths between them compare well with those of DNA-dependent DNA polymerases. However, since this family is composed of four members with only two sufficiently distant ones, more sequences are needed in order to narrow the number of conserved regions that are truly functionally essential.

RNA-dependent polymerases

Of the four motifs of viral RNA-dependent RNA polymerases, as well as reverse transcriptase sequences recently identified by Poch et al. (1989), two match motifs A and C of DNA-dependent polymerases. (These motifs are the only sequence features shared by all these sequences.) The 31% identity alignment of the hepatitis B reverse transcriptase from Woodchuck (Galibert et al., 1982) with the Herpes simplex virus DNA pol α (Gibbs et al., 1985) strongly supports the relationships between DNA- and RNA-dependent polymerases. The closeness of these two sequences is also sensible on evolutionary grounds, because reverse transcriptases have already been postulated to be an essential intermediary step in going from an RNA-dependent RNA polymerase to a DNA-dependent DNA polymerase (Lazcano et al., 1988). Furthermore, it should be pointed out that hepatitis B viruses are the only viruses encoding reverse transcriptase activity that have a DNA genome.

A site-directed mutagenesis experiment has pointed to the catalytic importance of the Asp–Asp doublet in motif C of RNA-dependent polymerases (Inokuchi and Hirashima, 1987). In addition, site-specific mutagenesis of the human immunodeficiency virus reverse transcriptase revealed that mutation of the first conserved Asp of motif C into His completely destroyed activity (Larder et al., 1987b). The mutation of the other strictly conserved Asp residue in motif A into Glu had a similar effect.

The sequence length between motifs A and C in the RNA-dependent polymerases averages ~30 residues, which is considerably shorter than the 130–170 amino acids in the pol Is, though one sequence from influenza virus contains 118 spacer residues, comparable with some pol αs (Figure 3). Once again, helix P and part of the helix Q and strands 10 and 11 are possible candidates for deletion. The shortest of all the distances (38 residues) suggests that deletions can occur in the M and N helices and in the loop following helix O, in that family.

Perspectives and conclusion

Argos (1988) previously discussed the conservation of motif C in pol αs and RNA-dependent polymerases; we extend this observation to the pol Is, pol βs and some DNA-dependent RNA polymerases. Apparently the fully conserved first Asp of the Asp–Asp doublet of motif C is catalytically more important than the second. We also contend that there is at least another motif upstream (motif A), with another strictly conserved aspartate located in a turn between a beta strand and an alpha helix, in spatial proximity of motif C; genetic experiments suggest that this motif might also be involved in the catalytic process. One of the many possible functional roles for these residues is that the strictly conserved Asp residue in motif C may co-operate with the one of motif A to bind a magnesium ion that would be part of the dNTP binding site. This site would be located at the bottom of a cleft containing the DNA. This cleft, similar to the one of the Klenow fragment, is already apparent in a 4 Å map of [illegible] DNA-dependent RNA polymerases (Stoa et al., 1989). If this is true, the different families of polymerases could be the result of divergence from a common ancestor, with essentially the same folding; or the conservation of conserved residues in the same spatial arrangement in the catalytic site could be due to convergent evolution, as observed in the subtilisin and chymotrypsin active site.

In spite of wide apparent sequence variability (likely to reflect a very ancient divergence), it is very tempting to adopt the strong, structurally unifying principle stating that many polymerases may fold like the known tertiary architecture of the *E.coli* pol I. Consistent with this hypothesis is the experimental observation that it is possible to change the template or substrate specificity of certain polymerase types if Mg^{2+} is replaced by Mn^{2+} (see Lazcano et al., 1988). This model, however, should be viewed as speculative, even though several lines of evidence point to this unifying conclusion. We believe it deserves attention, because site-directed mutagenesis experiments aimed at probing the catalytic site and template specificity should benefit from our hypothesis, which gives a possible structural framework for

future experiments. If reported crystals of a heterodimer of the reverse transcriptase for human immunodeficiency virus prove to be diffraction worthy (Lowe et al., 1988) and if the high resolution structure of T7 RNA polymerase soon becomes available, then the final test for the hypothesis made here should be forthcoming.

References

[illegible]

Received on November 28, 1989; accepted on February 28, 1990

EXHIBIT B

Docket No.: NEB-166-PUS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS: Jack et al. EXAMINER: Hutson

SERIAL NO.: 10/089,027 ART UNIT: 1652

DATE FILED: March 26, 2002

TITLE: Incorporation of Modified Nucleotides By Archaeon DNA Polymerases And Related Methods

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION UNDER 37 C.F.R. §1.131

As a below named inventor, I hereby declare that:

1. My name is Dr. William Jack, Research Director for the DNA Enzymes Division at New England Biolabs Inc. My resume is attached.

2. I have been studying the structure and function of DNA polymerases for over 16 years.

3. I was a member of the group of scientists at New England Biolabs that isolated, characterized, and cloned the first hyperthermophilic archaeal DNA polymerase. Our continuing work with archaeon DNA polymerases identified a surprisingly homogeneous set of enzymes. We claimed this group of DNA polymerases in US Patent 5,500,363. In this patent, the United States Patent and Trademark Office recognized the validity of our claim to a class of archaeon DNA polymerases defined by the DNA encoding the enzyme and its

ability to hybridize under defined conditions to various specified DNA sequences. The group was exemplified by T.litoralis (Vent), GBD (Deep Vent), and 9°N DNA Polymerases.

4. We also found that this group of polymerases had a high degree of amino acid sequence identity. A comparative three-dimensional alignment of members of this group of enzymes showed a high degree of structural conservation, consistent with the observed high degree of primary amino acid sequence identity/similarity. See for example, Vent (Rodriguez, et al., 2000), Tgo (Hopfner, et al., 1999), D. Tok (Zhao, et al., 1999), and KOD (Hashimoto, et al., 2001) DNA Polymerases.

5. The structural equivalence of this group of polymerases is further supported by experiments reported in Example 10 of the above application in which we show that mutation of an analogous residue in Vent and 9°N DNA Polymerases yields enzymes with equivalent acyclonucleotide incorporation efficiencies.

6. We discovered that this group of enzymes is capable of efficiently utilizing acyclonucleotides as substrates. We demonstrated this property using four examples of polymerases within this tightly defined group. Any molecular biologist of ordinary skill in the art would expect from these findings that this property would occur in all members of the enzyme group defined above.

7. Additionally, my colleagues and I have published articles in peer reviewed journals discussing the physical basis for the preferential incorporation of acyclonucleotides, and also for the enhanced incorporation with Vent A488L and 9°N A485L DNA Polymerase mutants. See Gardner, et al. (2004) on page 11841, column 1, paragraph 2 and page 11841, column 2, paragraph

1, respectively.

8. I assert that the combination of the high degree of homogeneity in DNA and amino acid sequences of archaeon DNA polymerases, plus the structural evidence that modification of specific amino acids alters enzyme specificity, would be sufficient to assure a person of ordinary skill in the art that the class of polymerases as defined above will interact with acyclonucleotide substrates as shown in the above application.

9. To further support the above statements, we have conducted additional experiments to confirm that archeon Family B polymerases with an amino acid sequence identity of greater than 30% can utilize acyclonucleotides as a substrate. This data is attached to the present declaration as appendix 1.

9. I further declare under penalty of perjury pursuant to laws of the United States of America that the foregoing is true and correct and that the Declaration was executed by me on:

William E. Jack

Date: 4 May 2006

References:

A. Gardner, C.M. Joyce, W.E. Jack (2004) J. Biol. Chem. 279, 11834-11842.

H. Hashimoto, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka, T. Inoue, Y. Kai (2001) J. Mol. Biol. 306, 469-477.

K.P. Hopfner, A. Eichinger, R.A. Engh, F. Laue, W. Ankenbauer, R. Huber, B. Angerer (1999) Proc. Natl. Acad. Sci. USA 96, 3600-3605.

A.C. Rodriguez, H.W. Park, D. Mao, L.S. Beese (2000) J. Mol. Biol. 299, 447-462.

Y. Zhao, D. Jeruzalmi, J. Moarefi, L. Leighton, R. Lasken, J. Kuriyan (1999) Structure 7, 1189-1199.

William Erle Jack
New England Biolabs
240 County Road
Ipswich, MA 01938
(978) 380-7257
(978) 921-1350 (fax)
email: jack@neb.com

RESEARCH INTERESTS

Enzymatic and structural aspects of protein-nucleic acid interactions. Thermostable DNA polymerase kinetics and function.

RESEARCH EXPERIENCE

New England Biolabs (Beverly, MA).

2005-present	Division Head, DNA Enzymes
1987-present	Senior Staff Scientist Research: Kinetic characterization of thermostable DNA polymerases. Creation and characterization of DNA polymerase variants with altered substrate recognition. Over-expression and characterization of restriction and modification enzymes.
2000-present	New England Biolabs Institutional Biosafety Committee Chair

Rockefeller University (NY, NY) Laboratory of Biochemistry and Molecular Biology.

1983-1987	Postdoctoral Fellow in the laboratory of R.G. Roeder. Research: Structural and functional characterization of wild type and mutant forms of *Xenopus* RNA polymerase III transcription factor A. Glucocorticoid hormone-induced transcription enhancement *in vitro*.

Duke University (Durham, NC) Department of Biochemistry.

1977-1983	Graduate Student in the laboratory of P. Modrich. Research: Kinetics and thermodynamics of DNA site location, recognition and cleavage by *Eco*RI endonuclease.

EDUCATION

Doctor of Philosophy (Biochemistry), Duke University, 1983 (Paul Modrich, advisor).
Bachelor of Arts (Chemistry), *Magna Cum Laude*, University of Utah, 1977.

TRAINING

2000	Sixth National Symposium on Biosafety: Prudent Practices for the New Millennium (Conducted by the Centers for Disease Control and Prevention)

PUBLICATIONS

"Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase," Andrew F. Gardner and **William E. Jack**, *J. Biol. Chem.* 279, 11834-11842 (2004).

"Acyclic and Dideoxy Terminator Preferences Denote Divergent Sugar Recognition by Archaeon and Taq DNA Polymerases," Andrew F. Gardner and **William E. Jack**, *Nucleic Acids Res.*, 30, 605-613 (2002).

"The Kinetic Mechanism of *Eco*RI Endonuclease," David J. Wright, **William E. Jack** and Paul Modrich, *J. Biol. Chem.* 274, 31896-31902 (1999).

"Determinants of Nucleotide Sugar Recognition in an Archaeon DNA Polymerase," Andrew F. Gardner and **William E. Jack**, *Nucleic Acids Res.* 27, 2545-2553 (1999).

"Photochemically Initiated Protein Splicing", Sandra N. Cook, **William E. Jack**, Xiaofeng Xiong, Lora E. Dudley, Jonathan A. Ellman, Peter G. Shultz, and Christopher J. Noren, *Angew. Chem. Int. Ed. Engl.* **34**, 1628-1630 (1995).

"Three-dimensional Structure of the Adenine-specific DNA Methyltransferase M·TaqI in Complex with the Cofactor S-adenosylmethionine", J. Labahn, J. Granzin, G. Schluckebier, D.P. Robinson, **W.E. Jack**, I. Schildkraut, and W. Saenger, *Proc. Natl. Acad. Sci. USA* **91**, 10957-10961 (1994).

"Expression, Purification, and Crystallization of Restriction Endonuclease *Pvu*II With DNA Containing Its Recognition Site", K. Balendiran, Joseph Bonventre, Roger Knott, **William Jack**, Jack Benner, Ira Schildkraut, and John E. Anderson, *Proteins: Structure, Function, and Genetics* **19**, 77-79 (1994).

"Protein Splicing Elements: Inteins and Exteins-A Definition of Terms and Recommended Nomenclature", Francine B. Perler, Elaine O. Davis, Gary E. Dean, Frederick S. Gimble, **William E. Jack**, Norma Neff, Christopher J. Noren, Jeremy Thorner and Marlene Belfort, *Nucleic Acids Res.* **22**, 1125-1127 (1994).

"Characterization of a DNA Polymerase from the Hyperthermophile Archaea *Thermococcus litoralis*", Huimin Kong, Rebecca B. Kucera and **William E. Jack**, *J. Biol. Chem.*, **268**, 1965-1975 (1993).

"Protein Splicing Removes Intervening Sequences in an Archaea DNA Polymerase", Robert A. Hodges, Francine B. Perler, Christopher J. Noren and **William E. Jack**, *Nucleic Acids Res.* **20**, 6153-6157 (1992).

"Intervening Sequences in an Archaea DNA Polymerase Gene", Francine B. Perler, Donald G. Comb, **William E. Jack**, Laurie S. Moran, Boqin Qiang, Rebecca B. Kucera, Jack Benner, Barton E. Slatko, Donald O. Nwankwo, S. Kay Hempstead, Clotilde K.S. Carlow and Holger Jannasch, *Proc. Natl. Acad Sci. USA* **89**, 5577-5581 (1992).

"Overexpression, Purification and Crystallization of *Bam*HI Endonuclease", **William E. Jack**, Lucia Greenough, Lydia F. Dorner, Shuang-yong Xu, Teresa Strzelecka, Aneel K. Aggarwal and Ira Schildkraut, *Nucl. Acids Res.* **19**, 1823-1829 (1991).

"Nucleotide Sequence of the *FokI* Restriction-Modification System: Separate Strand-Specificity Domains in the Methyltransferase", Mary C. Looney, Laurie S. Moran, **William E. Jack**, George R. Fechery, Jack S. Benner, Barton E. Slatko and Geoffrey G. Wilson, Gene **80**, 193-208 (1989).

"M.FokI methylates adenine in both strands of its asymmetric recognition sequence", David Landry, Mary C. Looney, George R. Fechery, Barton E. Slatko, **William E. Jack**, Ira Schildkraut, and Geoffrey G. Wilson, *Gene* **77**, 1-10.

"Mechanism of specific site location and DNA cleavage by *EcoRI* endonuclease", Brian J. Terry, **William E. Jack** and Paul Modrich, *Gene Amplif Anal.* volume 5, pp. 103-118 (1987)

"Facilitated diffusion during catalysis by EcoRI endonuclease. Nonspecific interactions in EcoRI catalysis", Brian J. Terry, **William E. Jack** and Paul Modrich, *J Biol. Chem.* **260**: 13130-13137 (1985).

Participation of Outside DNA sequences in the EcoRI Endonuclease Reaction Pathway, **William E. Jack**, dissertation (Duke University, Durham, NC) (1983).

"Thermodynamic parameters governing interaction of EcoRI endonuclease with specific and nonspecific DNA sequences", Brian J. Terry, **William E. Jack**, Robert A. Rubin and Paul Modrich, *J Biol. Chem.* , **258**, 9820-9825 (1983).

"Involvement of outside DNA sequences in the major kinetic path by which EcoRI endonuclease locates and leaves its recognition sequence", **William E. Jack**, Brian J. Terry, and Paul Modrich, *Proc Natl. Acad Sci U S A* 79: 4010-4014 (1982).

"DNA determinants important in sequence recognition by *EcoRI* endonuclease", A-Lien Lu, **William E. Jack** and Paul Modrich, *J Biol. Chem.* **256**: 13299-13306 (1981).

"Structures and Mechanisms of *EcoRI* DNA Restriction and Modification Enzymes", **William E. Jack**, Robert A. Rubin, Andrea Newman and Paul Modrich, *Gene Amplif. Anal.* volume 1, p 165-179 (1981).

US PATENTS

"Use of site-specific nicking endonucleases to create single-stranded regions and applications thereof," **William E. Jack**, Ira Schildkraut, Julie Forney Mevis, US Patent 6,660,475, December 9, 2003.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,834,285, November 10, 1998.

"Modified proteins comprising controllable intervening protein sequences or their elements methods of producing same and methods for purification of a target protein comprised by a modified protein," Donald G. Comb, Francine B. Perler, **William E. Jack**, Ming-Qun Xu, Robert A. Hodges, Christopher J. Noren, Shaorong S.C. Chong, Eric Adam, Maurice Southworth, US Patent 5,834,247, November 10, 1998.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,500,363, March 19, 1996.

"Modification of protein by use of a controllable intervening protein sequence," Donald G. Comb, Francine B. Perler, **William E. Jack**, Ming-Qun Xu, Robert A. Hodges, US Patent 5,496,714, March 5, 1996.

"Recombinant thermostable DNA polymerase from archaebacteria," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,352,778, October 4, 1994.

"Purified thermostable DNA polymerase obtainable from thermococcus litoralis," Donald G. Comb, Francine Perler, Rebecca Kucera, **William E. Jack**, US Patent 5,322,785, June 21, 1994.

doi:10.1006/jmbi.2000.3728 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (2000) 299, 447–462

JMB

Crystal Structure of a Pol α Family DNA Polymerase from the Hyperthermophilic Archaeon *Thermococcus sp.* 9°N-7

A. Chapin Rodriguez†, Hee-Won Park†, Chen Mao and Lorena S. Beese*

Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

*Corresponding author

The 2.25 Å resolution crystal structure of a pol α family (family B) DNA polymerase from the hyperthermophilic marine archaeon *Thermococcus* sp. 9°N-7 (9°N-7 pol) provides new insight into the mechanism of pol α family polymerases that include essentially all of the eukaryotic replicative and viral DNA polymerases. The structure is folded into NH_2-terminal, editing 3′-5′ exonuclease, and polymerase domains that are topologically similar to the two other known pol α family structures bacteriophage RB69 and the recently determined *Thermococcus gorgonarius*, but differ in their relative orientations and conformations.

The 9°N-7 polymerase domain structure is reminiscent of the "closed" conformation characteristic of ternary complexes of the pol I polymerase family obtained in the presence of their dNTP and DNA substrates. In the apo-9°N-7 structure, this conformation appears to be stabilized by an ion pair. Thus far, the other apo-pol α structures that have been determined adopt open conformations. These results therefore suggest that the pol α polymerases undergo a series of conformational transitions during the catalytic cycle similar to those proposed for the pol I family. Furthermore, interaction of the orientations of the fingers and exonuclease (sub)domains relative to the palm subdomain that contains the pol active site suggests that the exonuclease domain and the fingers subdomain of the polymerase can move as a unit and may do so as part of the catalytic cycle. This provides a possible structural explanation for the interdependence of polymerization and editing exonuclease activities unique to pol α family polymerases.

We suggest that the NH_2-terminal domain of 9°N-7 pol may be structurally related to an RNA-binding motif, which appears to be conserved among archaeal polymerases. The presence of such a putative RNA-binding domain suggests a mechanism for the observed autoregulation of bacteriophage T4 DNA polymerase synthesis by binding to its own mRNA. Furthermore, conservation of this domain could indicate that such regulation of pol expression may be a characteristic of archaea. Comparison of the 9°N-7 pol structure to its mesostable homolog from bacteriophage RB69 suggests that thermostability is achieved by shortening loops, forming two disulfide bridges, and increasing electrostatic interactions at subdomain interfaces.

Keywords: Archea; X-ray structure; replication; exonuclease; family B DNA polymerase

Introduction

DNA polymerases catalyze the template-directed addition of nucleotides onto the 3′-OH group of the DNA primer terminus. These enzymes replicate DNA with the required accuracy essential for genetic

†Contributed equally to the manuscript.
Abbreviations used: pol, polymerase; Tgo, *Thermococcus gorgonarius*; ddNTP, dideoxynucleotides.
E-mail address of the corresponding author: sbl@biochem.duke.edu

0022-2836/00/030447–16 $35.00/0

mic stability, but generate sufficient mutations to stimulate and maintain evolution. Unlike Eucarya and Bacteria, relatively little is known about DNA replication in Archaea (Forterre et al., 1996), one of the three major evolutionary lineages of life (Woese et al., 1990). Archaea play a significant role in the biosphere, accounting for up to 30% of the biomass in certain Antarctic waters (De Long et al., 1994), and exhibit much greater diversity than had originally been suspected (Barns et al., 1996). Many characterized archaeal species are adapted to live in environments of extreme temperature, pressure, salinity, and/or pH such as hydrothermal vents, and hot springs (Rees & Adams, 1995).

Although archaeal cells share many morphological features with Bacteria, archaeal proteins involved in gene expression including DNA replication, transcription, and translation have been found to be similar to those from Eucarya (Edgell & Doolittle, 1997; Bult et al., 1996). In particular, most of the archaeal DNA polymerases that have been sequenced belong to the α-like polymerase family (family B) that includes essentially all the eukaryotic replicative and viral DNA pols (Braithwaite & Ito, 1993; Edgell et al., 1997).

Crystal structures exist for DNA pols from each of four families: pol I (family A), pol α (family B), pol β (family X) and reverse transcriptase (reviewed by Joyce & Steitz, 1994; Doublié et al., 1999). Although pols from different families are structurally quite diverse, several common features have emerged. The pol domain from each resembles a right hand and may be further divided into palm, fingers, and thumb subdomains, as was originally described for the large fragment of *Escherichia coli* pol I (Klenow fragment) (Ollis et al., 1985). All polymerases appear to share the same mechanism for nucleotidyl transfer involving two divalent metal ions (reviewed by Beese & Steitz, 1995). In addition, based on structures containing DNA and dNTP bound to pols from pol I, pol β, and reverse transcriptase families, a conformational change in the fingers subdomain from an open to a closed conformation is proposed to occur during the catalytic cycle (reviewed by Doublié et al., 1999).

The pol α family polymerases are of medical importance as targets for development of antiviral and anticancer therapeutics. For example, human pol α is a target in the treatment of acute myelogenous leukemia and chronic lymphocytic leukemia (Keating et al., 1982; Robertson & Plunkett, 1996) and a variety of nucleotide analogs with antitumor activity inhibit strand elongation by pol α (Huang & Plunkett, 1995; Condit & Maizel, 1996). Furthermore, polymerases, particularly those that are thermostable, have a number of critical biotechnological applications ranging from PCR to cloning and DNA sequencing. Despite their biological, medical and biotechnological importance, the pol α class of polymerases has not been structurally as well characterized as other DNA polymerase families.

Here we report the 2.25 Å resolution crystal structure of a pol α family DNA polymerase from the hyperthermophilic marine archaeon *Thermococcus* sp. 9°N-7 (9°N-7 pol). *Thermococcus* sp. 9°N-7 was isolated from a hydrothermal vent at 9° N latitude off the East Pacific Rise (Southworth et al., 1996). The structure is folded into NH$_2$-terminal, editing 3′-5′ exonuclease, and polymerase domains that are topologically similar to the two other known pol α family structures (bacteriophage RB69 (Wang et al., 1997) and the recently determined *Thermococcus gorgonarius* (Tgo) (Hopfner et al., 1999), but differs in their relative orientation and conformation.

The pol domain structure is reminiscent of the "closed" conformation characteristic of ternary complexes of the pol I polymerase family obtained in the presence of their dNTP and DNA substrates. In the apo-9°N-7 structure, this conformation appears to be stabilized by an ion pair. Thus far, the two other apo-pol α structures that have been determined adopt open conformations. These results therefore suggest that the pol α polymerases undergo a series of conformational transitions during the catalytic cycle similar to those proposed for the pol I family. Furthermore, comparison of the orientations of the fingers and exonuclease domains relative to the palm subdomain that contains the pol active site suggests that the exonuclease domain and the fingers subdomain of the polymerase can move as a unit, and may do so as part of the catalytic cycle. This provides a possible structural explanation for the interdependence of polymerization and editing exonuclease activities unique to pol α family polymerases.

We suggest that the NH$_2$-terminal domain of 9°N-7 pol is structurally homologous to the ββββαββ RNA-binding motif with an exposed patch of aromatic amino acid residues. Bacteriophage T4 DNA pol, which is homologous to 9°N-7 pol, is known to bind its own mRNA and repress its own synthesis. The homology relationships to the RNA-binding motif suggest a structural basis for this regulatory mechanism. Furthermore, the conservation of this domain in other archaeal pols suggests that such autogenous regulation of pol expression may be general for archaea.

Results and Discussion

Crystal structure of *Thermococcus* sp. 9°N-7 pol

The structure of the full-length, 775-residue enzyme (bearing the double mutation D141A and D143A) was determined using the multiple isomorphous replacement method to a resolution of 2.25 Å. The current model has an *R*-factor of 23.9% (R_{free} = 30.8%) (Table 1). A Ramachandran plot of the model shows 86.8% of the residues in the most favored region and the remainder in additional allowed regions (12.4%) and generously allowed regions (0.8%). A total of 37 residues are not traced in the model and lie in regions of poorly defined electron density. The first of these gaps

Table 1. Crystallographic data collection and refinement statistics

	Native			Derivative						
Data	NAT-1[a]	NAT-2	NAT-3	THI-1[a]	THI-2	PtCl-1[a]	PtCl-2	PIP-1[a]	PIP-2	BAHg[a]
A. Cell dimensions (Å) $P2_12_12_1$										
a	96.1	94.8	95.3	96.7	96.2	95.8	96.3	97.0	95.4	96.1
b	101.1	98.2	98.7	100.7	99.6	100.8	99.9	100.5	99.6	100.2
c	113.3	112.3	112.6	113.3	113.2	113.9	112.9	111.6	112.3	113.1
Resolution (Å)	25-3.0	25-2.3	25-2.1	25-3.3	25-2.8	25-3.5	25-3.5	25-4.5	25-2.8	25-2.5
Highest resolution bin (Å)	3.06-3.00	2.34-2.30	2.15-2.10	3.36-3.30	2.85-2.80	3.56-3.50	3.56-3.50	4.57-4.50	2.85-2.80	2.56-2.50
Completeness[b] (%)	99.0 (91.7)	99.2 (49.0)	94.2 (92.5)	99.8 (99.9)	90.3 (90.9)	98.6 (99.9)	97.3 (64.5)	99.6 (99.6)	96.6 (82.8)	97.9 (96.4)
Unique reflections	22,082	62,987	58,966	14,213	23,885	14,021	13,616	9990	25,788	33,769
Total reflections	164,954	552,366	462,420	177,293	575,987	177,379	119,366	230,698	255,831	230,642
R_{merge}[c]	8.7	5.1	5.3	9.1	10.2	7.3	11.7	11.3	14.6	8.8
R_{iso}[d]				27.7	22.9	39.6	30.3	20.2	15.1	36.0
Phasing power[e]				0.9	2.1	1.6	1.2	2.1	6.8 (to 4 Å)	1.8

B. Refinement (F > 2σF)

Resolution limits (Å)	25-2.25	rmsd from ideality		
Highest resolution bin (Å)	2.35-2.25		in bond length (Å)	0.009
Reflections	55,922		in bond angle (deg.)	1.280
Completeness[b] (%)	95.3 (86.9)			
R-factor[f] (%)	21.9		Non-hydrogen atoms	5,996
R_{free}[f]	30.8		Water molecules	169

[a] Abbreviations: NAT, native; THI, sodium ethylmercurithiosalicylate; PtCl, K_2PtCl_4; PIP, di-μ-iodobis(ethylenediamine)diplatinum (II) nitrate; BAHg, Baker's mercurial.

[b] Numbers in parentheses correspond to completeness in the highest-resolution bin.

[c] $R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma I$, where $\langle I\rangle$ refers to the average intensity I of multiple measurements of the same reflection.

[d] $R_{iso} = \Sigma|F_{native} - F_{deriv}|/\Sigma F_{native}$, where "native" and "deriv" refer to the native and derivative data.

[e] Phasing power $= \langle F_H\rangle/\langle E\rangle$, where $E^2 = \Sigma(F_{PH,obs} - F_{PH,calc})^2$ and F_{PH} and F_H are the protein derivative and heavy atom structure factors, respectively.

[f] R-factor and $R_{free} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$. R_{free} was calculated over 5% of the structure factor amplitudes not used in the refinement.

occurs at the bottom of the palm domain (residues 568-575), and the remainder are within the thumb region that is frequently observed to be partially disordered in apo polymerase structures, as is also the case here (e.g. Ollis et al., 1985; Kiefer et al., 1997). Although no disulfide bridges were included in the refinement, four Cys residues showed anomalous peaks in a difference Fourier map and side-chain distances and angles consistent with two disulfide bridges (Cys428:442, Cys506:509).

The structure of 9°N-7 pol reveals features common to all DNA pol structures as well as those that may be unique to archaeal pols. The overall shape of the enzyme can be described as a disc with a central hole that is folded into NH_2-terminal, 3'-5' exonuclease, and polymerase domains (Figure 1(a) and (b)). Like all other pols of known structure, the pol domain resembles a right hand and may be further divided into palm, fingers, and thumb subdomains, as was originally described for the large fragment of E. coli pol I (Klenow fragment) (Ollis et al., 1985). 9°N-7 pol is similar in structure to the pol α family polymerase from the mesophilic bacteriophage RB69 (RB69 pol) (Wang et al., 1997); although a number of these (sub)domains are shorter than in RB69 pol (Figure 1(c)). Nearly all these sequence length differences are attributable to loop segments that are fewer and shorter in the hyperthermostable 9°N-7. As was first observed in the RB69 pol structure (Wang et al., 1997), the 3'-5' exonuclease domain lies on the opposite side of the palm in comparison to pol I family polymerases. This domain arrangement is also seen in 9°N-7 pol and in Tgo pol (Hopfner et al., 1999), indicating that this result is likely to be general for the pol α family. The structural similarity between 9°N-7 and RB69 pols is significant given the low sequence identity (≤20%) in all but the active-site (palm) region, where sequence identity is 42% (Figure 2). Similar results hold for sequence alignments between 9°N-7 and human pol α.

NH_2-terminal domain

Many of the members of the pol α polymerase family, including archaeal pols, bacteriophage T4 and RB69 DNA pols, have an NH_2-terminal domain that is not observed in the pol I family. T4 pol is known to control its synthesis in vivo by a mechanism of autogenous regulation (Tuerk et al., 1990). The mRNA-binding activity has been located to within the first 108 residues of the pol (Wang et al., 1996), but the structure of a fragment comprising residues 1-388 of T4 pol failed to suggest a structural basis for RNA binding (Wang et al., 1996). Here, we note that certain structural similarities between the homologous region in the 9°N-7 pol and the U1A RNA-binding protein may provide a rationale for RNA binding by T4 pol.

The NH_2-terminal domain of 9°N-7 pol can be considered as three modules based on compactness of folding (Figure 3(a)). The first module comprises residues 1-33, a three-stranded β-sheet that interacts extensively with the 3'-5' exonuclease domain via predominantly electrostatic interactions. Residues 33-36 act as a flexible linker connecting the first module to the second (residues 37-123). The third module comprises residues 338-377.

The second module is folded into a βαββαβ motif, with two short β-strands, 5 and 6, inserted between the second and third elements. This motif occurs in a variety of proteins, and forms the basis for the most prevalent RNA binding motif, the RNA recognition motif (RRM). The RRM is present in the RNA-binding domains of hnRNP A1, spliceosomal protein U1A and U2B", and the sex lethal protein (Bard & Dreyfuss, 1994). Although an alignment of the NH_2-terminal domains of archaeal pols (Figure 3(b)), together with T4 and RB69 pols, shows that they lack the RNP1 and RNP2 sequence motifs that characterize the RRM (Bard & Dreyfuss, 1994), a number of highly conserved and invariant residues nevertheless emerges. Most of these residues fall in a cluster on the surface of the NH_2-terminal domains of 9°N-7 and RB69 pols which therefore could mark the location of an RNA binding site atop the β-sheet platform on the face away from helix A (Figure 3(c)).

Both a sequence alignment (Figure 3(b)) and a structural comparison (Figure 3(c)) reveal that T4 and RB69 pols lack helix A and strand 3 of the βαββαβ motif, perhaps explaining why no suggestive structural homologies to RNA-binding folds could be identified (Wang et al., 1996, 1997).

Experiments are needed to determine whether the NH_2-terminal domain of 9°N-7 pol binds RNA. Although the βαββαβ motif occurs in proteins that are not thought to interact with RNA (Bard & Dreyfuss, 1994), we find its presence in the NH_2-terminal domain of 9°N-7 pol, in a region known to bind RNA in T4 pol (Wang et al., 1996), to be highly suggestive of this. RNA-binding capability could hold for other archaeal pols as well, since sequence alignment of NH_2-terminal domains (Figure 3(b)) suggests that they share the βαββαβ motif.

We further speculate that just as T4 pol binds its mRNA to down-regulate its own synthesis, such autogenous regulation of pol expression might occur in archaea. Autogenous gene regulation is well documented in bacteria, and has at least one precedent in archaea. It has been identified in the synthesis of the MvaL1 ribosomal protein of *Methanococcus vannielii* (Hanner et al., 1994), and postulated for a ribosomal gene cluster from the halophile *Halobacterium cutirubrum* (Shimmin & Dennis, 1989). It is interesting that there is no structural evidence that such regulation extends to eukaryotes, as human pol α shows no significant sequence homology to the NH_2-terminal sequences aligned in Figure 3(b).

3'-5' Exonuclease domain

This domain is responsible for binding single-stranded DNA and excising mismatched bases in the elongated primer strand. The structure

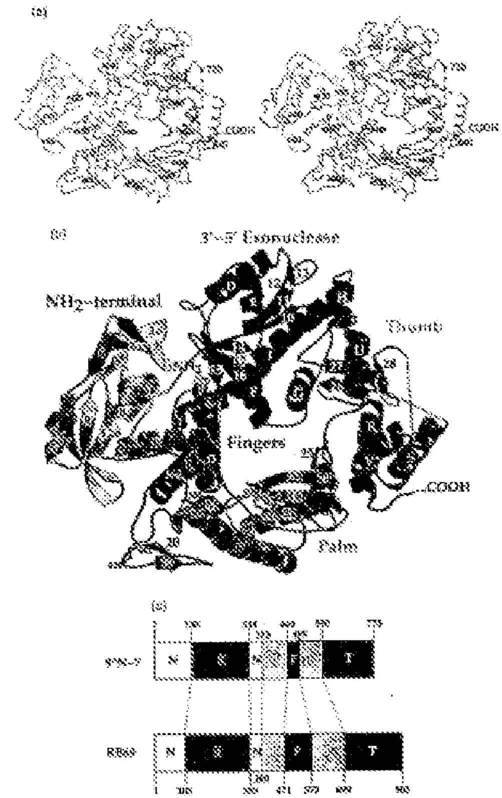


Figure 3. Structure of the Thermococcus sp. 9°N-7 DNA polymerase. The NH_2-terminal and 3'-5' exonuclease domains are colored yellow and green, respectively. The polymerase domain is divided into palm (brown), thumb (red), and fingers (blue) subdomains. Three highly conserved carboxylate groups (D404, D540, D542) mark the polymerase active site. (a) Stereoview of the C^{α} trace. Every 40th C^{α} is numbered. Broken lines indicate disordered regions of the protein. (b) Ribbon diagram with secondary structure elements defined according to DSSP (Kabsch & Sander, 1983). NH_2-terminal domain: 1, 1-10; 2, 18-23; 3, 33-35; 4, 37-42; A, 48-51; 5, 55-58; 6, 61-64; 7, 67-75; 8, 78-86; B, 92-101; 9, 106-110; C, 118-125; J, 341-344; K, 359-365. 3'-5' exonuclease domain: 10, 137-144; 11, 157-163; 12, 168-173; 13, 181-183; D, 187-201; 14, 205-208; E, 215-225; 15, 240-244; 16, 247-251; 17, 256-259; F, 260-266; G, 275-282; H, 293-300; I, 305-337. Polymerase domain: L, 376-379; 18, 397-404; M, 408-415; 19, 421-453; 20, 440-442; N, 448-466; O, 475-498; P, 507-532; 21, 538-539; 22, 543-547; Q, 553-567; 23, 578-580; 24, 583-598; 25, 600-606; R, 617-625; S, 636-651 (646-653 disordered); T, 657-660 (disordered); 26, 662-665; U, 677-688; 27, 698-703; 28, 714-716; V, 721-734; W, 742-756. (c) Schematic comparing the (sub)domains of Thermococcus sp. 9°N-7 and bacteriophage RB69 DNA polymerases. The domain boundaries for 9°N-7 pol were determined based upon a structure-based sequence alignment with RB69 pol (Figure 7) as defined for the RB69 pol (Wang et al., 1997).

USSN 10/089,027

Atty. Docket No. 2007651-0001

4567604v1

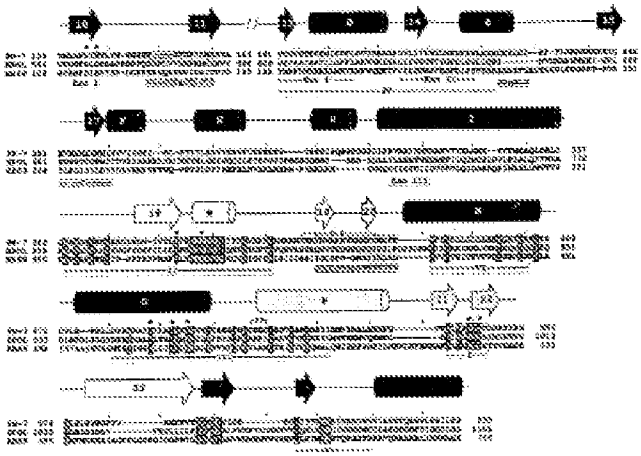


Figure 2. A three-way partial sequence alignment of *Thermococcus* sp. 9°N-7 pol (9N-7), RB69 pol (RB69), and human pol α (hPOLA). Dashes indicate gaps in the alignment, and segments not aligned are represented as amino acid residue spans within brackets. Ticks mark every 50 spaces. The 9°N-7 and RB69 pol alignment is based upon the crystal structures. The HPOL and RB69 alignment is from Wang *et al.* (1997), except for a few short segments assigned based upon the three sequences shown here. Indicated below the sequences and boxed in yellow are conserved motifs in the exonuclease (Blanco *et al.*, 1992) and polymerase (Wang *et al.*, 1997) domains. The secondary structure elements in 9°N-7 pol, as defined by DSSP, are given above the sequence. The structural elements are colored according to the scheme described in regard to Figure 1. Shown in purple in the 9°N-7 pol sequence are the archaeal polymerase motifs described by Edgell *et al.* (1997). Residues within the polymerase domain that are invariant in the three sequences (blue boxes; residues discussed in the section on dNTP binding, blue asterisk. The two disulfide bridges in the palm (C420-C442, C566-C708) are shown schematically.

reported here is that of a mutant of 9°N-7 pol lacking detectable exonuclease activity which was engineered to prevent degradation of DNA substrates during subsequent co-crystallization experiments. This 9°N-7exo⁻ pol was obtained by making two point mutations (D141A, E143A) in the Exo I (ExoI) motif highly conserved among the 3′-5′ exonuclease domains of many DNA pols (Kunkel, 1992; Blanco *et al.*, 1992). In the Klenow fragment (KF) of *E. coli* DNA pol I, these residues (D355, E357) are responsible for binding the catalytic metals and for hydrogen bonding with the 3′-OH of the terminal deoxynucleotide of the substrate DNA (Beese & Steitz, 1991).

Aside from loop segments that are shorter than those observed in RB69 pol (see below), the topology of the exonuclease domain in 9°N-7 pol is very similar to that of RB69 pol. The domains superimpose in the central β-sheet, containing the active site, with a root mean square deviation (rmsd) of 0.95 Å (39 Cᵅ atoms). The metal-binding residues not mutated in 9°N-7exo⁻ pol, D215 and D218, superimpose almost exactly on the corresponding RB69 pol residues (D323, D327).

It is now possible to assign a structural context to the four archaeal sequence motifs identified by Edgell *et al.* (1997). Three of the regions (A-C) lie within the exonuclease domain (Figure 2). Motif A forms part of the central β-sheet containing the active site; B, part of a solvent-exposed loop; and C, part of a five-stranded β-sheet nearly perpendicular to the central β-sheet. The fourth motif resides in the palm (see below).

Pol domain

This domain is responsible for the template-directed polymerization of dNTPs onto the growing primer strand of duplex DNA. Like other polymerases of known structure, the pol domain can be further divided into palm, fingers, and thumb subdomains. While the structure of the thumb of 9°N-7 and RB69 pols are highly similar, differences exist in the palm and fingers. Some of these differ-

ences correspond to features that appear unique to archaeal pols, while others support a hypothesis that a conformational change occurs in the fingers as part of the catalytic cycle.

Palm subdomain

The palm, which contains the active site for polymerization, shows a high degree of structural similarity to the palm subdomain of other DNA polymerases. It is as structurally similar to pol I family polymerases as to those of the pol α family. Its rms deviation from RB69 pol around the active site (blue region in Figure 4(b)) is 0.88 Å (26 Cα atoms). Together with the Tgo pol structure (Hopfner et al., 1999), this structure confirms for archaea the conservation of a common catalytic core. A significant difference between the palm subdomains in 9°N-7 and RB69 pols are the two disulfide bridges present in 9°N-7 pol, one joining Cys428 and 482 and another joining Cys506 and 509 (Figure 4(b)). Both the shortened loops and at least one disulfide bridge appear common to archaeal pols (see above). Indeed, the region containing one of the Cys residues in a disulfide bridge (C482) corresponds to the highly conserved archaeal motif D (Edgell et al., 1997; Figure 2). The Tgo pol structure shows the corresponding Cys residues to be "poised" for disulfide formation, but still in reduced form.

Until recently it was believed that all pols share a catalytic "triad" of carboxylate residues in the active site in the palm (Delarue et al., 1990). Wang et al. (1997) since recognized that only two of the carboxylate residues are invariant. The invariant carboxylates in 9°N-7 pol are D483 and D542. The third member of the triad, present as D580 in 9°N-7 pol, is not essential: mutations of the corresponding residue (D1082N) in human pol α retains catalytic function (Copeland et al., 1993). D580 in 9°N-7 pol may nevertheless be involved in binding the divalent metals required for catalysis. Mg^{2+} is normally the optimal metal for human pol α activity. The pol α D1082N mutant shows greater catalytic efficiency and fidelity with Mn^{2+} rather than Mg^{2+} (Copeland & Wang, 1993).

D580 in 9°N-7 pol interacts with the hydroxyl group of Y538 that is within hydrogen-bonding distance to D580. Substitution of this residue to Phe in human pol α (Y1026) causes only minor effects on catalysis but alters the pol metal affinity akin to the pol α D1082 mutation (Copeland & Wang, 1993). It seems likely that the hydroxyl moiety of Y538 in 9°N-7 pol helps to lock D580 in position for Mg^{2+}-specific binding. Consistent with this function is the strict conservation of Y538 among pol α family members (Braithwaite & Ito, 1993).

Fingers subdomain

The fingers subdomain of 9°N-7 differs in topology and relative conformation from RB69. The fingers of 9°N-7 pol are a simple helix-coil-helix, as in Tgo pol (Hopfner et al., 1999), whereas in the fingers of RB69 pol, the coil region is expanded with more secondary structure elements (Figures 2 and 5). The shorter fingers of 9°N-7 pol are conserved among the archaeal pols aligned by Edgell et al. (1997). It is possible that the fingers of archaeal pols define a minimal functional unit.

Different positions of the fingers subdomain relative to the palm are observed in the 9°N-7 and RB69 pol structures (Figure 3(a)). The fingers of Tgo pol (Hopfner et al., 1999) show a position intermediate between that in 9°N-7 and RB69 pols, when the palm subdomains of all three enzymes are aligned. It is interesting to note that the fingers subdomains of polymerases in the pol I family adopt different positions during the catalytic cycle (reviewed by Doublié et al., 1999). An open position corresponds to that seen in the apoenzyme form (Ollis et al., 1985; Kiefer et al., 1998; Kornetov et al., 1995; Kiefer et al., 1997) and the form bound to duplex DNA (Beese et al., 1993; Kiefer et al., 1998). A closed conformation has been observed in the ternary replication complexes of bacteriophage T7 pol (Doublié et al., 1998), and Klentaq (Li et al., 1998) with bound DNA and dNTP. An analogous conformational change has been observed in ternary complexes of human immunodeficiency virus reverse transcriptase (Huang et al., 1998) and rat pol β (Pelletier et al., 1994). In the closed conformation, the fingers reside towards the palm to form a binding pocket for dNTPs.

The differences in position of the fingers subdomain in the three pol α family crystal structures suggest that the fingers of pol α family pols move during catalysis, analogous to that observed for the other polymerase families. It is interesting to note that if this is the case, there must be a corresponding movement in the position of the 3′-5′ exonuclease domains not required in the other polymerase families as will be discussed below. If the position of the fingers in 9°N-7 pol more closely approximates a closed conformation, it is not clear why they would adopt a position previously observed only in ternary complexes with bound dNTP and DNA. The fingers of 9°N-7 pol may be stabilized in this conformation because of a salt-bridge between E378 in the palm and K487 on helix O of the fingers. These residues are highly conserved among archaeal pols (Edgell et al., 1997) and both pol I and pol α families (Braithwaite & Ito, 1993). The corresponding salt-bridge does not form in polymerases of the pol I family because the fingers helix O lies too far from the palm. The fingers of Tgo pol, in fact, are rotated slightly away from the active site, relative to 9°N-7 pol, such that the E378-K487 salt-bridge cannot form. Another possible explanation for the difference in finger positions are the disulfide bridges present in 9°N-7 pol but absent in the Tgo pol structure and in pol I family structures. At least one of the disulfides (Cys428-482) in 9°N-7 pol could be directly involved in orienting the fingers relative to the palm (Hopfner et al., 1999).

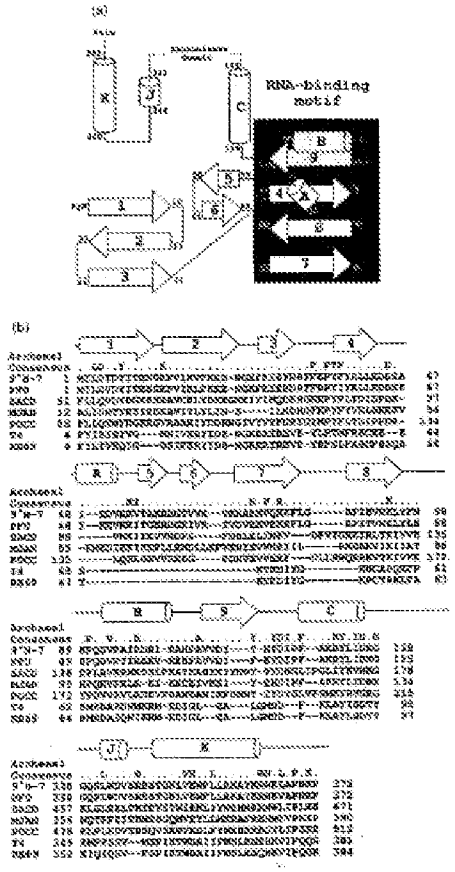


Figure 3 (legend opposite)

Model for DNA and dNTP binding

Based on the high degree of structural homology of the palm subdomains between 9°N-7 and pol I family pols, DNA and dNTP substrates from the bacteriophage T7 pol ternary complex (Doublie et al., 1998) were modeled into the 9°N-7 pol active site. The model shown in Figure 6 provides further

C 9°N-7 RB69

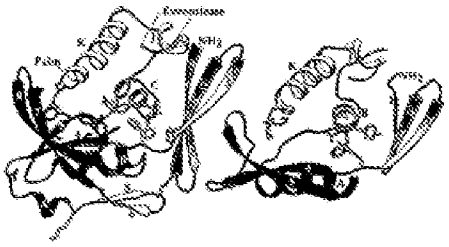


Figure 3. The RNA-binding motif in the NH_2-terminal domain of 9°N-7 pol. (a) Topology diagram of the complete NH_2-terminal domain (residues 1-429, 538-572). The RNA-binding motif βαββαβ known as the RNP consensus motif (Burd & Dreyfuss, 1994), is boxed. (b) Sequence alignment of the NH_2-terminal domains of 9°N-7 pol, RB69 pol, T4 pol, and archaeal pols. Alignment of 9°N-7 pol and T4/RB69 is based upon the crystal structures, and that of 9°N-7 pol and the other archaeal polymerases is based upon sequence alignments of 13 sequences among those considered by Edgell et al. (1997) (data not shown). The archaeal polymerase alignment was performed with the PILEUP algorithm in the GCG package (University of Wisconsin Genetic Computer Group). Secondary structure elements corresponding to 9°N-7 pol are given above the sequence. A consensus sequence was derived for the archaeal polymerases at those positions where at least 70% of the 13 sequences shared the same residue. Boxed in yellow are those residues conserved between the archaeal consensus and both bacteriophage (T4, RB69) sequences. Position 367 in 9°N-7 pol is starred (see the text for discussion). Abbreviations are as follows: PFU, *Pyrococcus furiosus*; SACO, *Sulfolobus solfataricus*; MJAN, *Methanococcus jannaschii*; PCKC, *Pyrodictium occultum* B1. (c) Ribbon representation of the NH_2-terminal domain of 9°N-7 (left) and RB69 (right) pols. Least-squares Cᵅ superposition was performed over the region of 9°N-7 pol including strand 6, part of strand 8, and helices B and C, and the domains were separated for side-by-side comparison. Shown in green is the βαββαβ RNA-binding motif. Charged and aromatic archaeal conservative residues are shown with green side-chains, and yellow side-chains correspond to the residues boxed in yellow in (b). The loop between β strands 7 and 8 in 9°N-7 pol corresponds to the conformationally variable loop 3 in the canonical RNP motif (Shamoo *et al.*, 1995).

evidence that the position of the fingers in 9°N-7 pol more closely approximates a closed conformation and their position in RB69 pol approximates an open conformation. This model of a ternary complex for a pol α family polymerase places the dNTP within hydrogen-bonding distance of residues on the fingers O helix that are highly conserved and known by mutagenesis to be functionally important. The corresponding residues on fingers helix P of the RB69 pol are farther away and cannot directly interact with dNTP.

The model places residues Y408 and Y494 near the deoxyribose moiety of the incoming dNTP. These residues appear to be functionally analogous to I888 and Y926 of T7 pol, which are responsible for discriminating between deoxy- and ribonucleotides (rNTPs). Y408 is invariant among the pol α family in the alignment by Braithwaite & Ito (1993) and nearly invariant (one exception) among archaeal pols aligned by Edgell et al. (1997). Mutation of the corresponding residue (Y412) to Val in an exonuclease-deficient *Thermococcus litoralis* (Vent) pol causes a 200-fold loss of discrimination against rNTPs. The aromatic ring appears to be the functionally important moiety, as mutating Y412 to Phe conserves wild-type discrimination levels (Gardner & Jack, 1999).

Y926 in T7 pol (F762 in Klenow fragment) has been dubbed the "release selectivity site" (Tabor & Richardson, 1995). A Phe residue at this position confers selectivity against incorporation of dideoxyribonucleotides (ddNTPs), whereas a Tyr residue in this position allows efficient incorporation of both nucleotide species. The presence of Tyr (Y494) in this position in 9°N-7 pol suggests the ability to incorporate dideoxynucleotides, as do Vent (Gardner & Jack, 1999) and human pol α (Copeland et al., 1992). In fact, Tyr is invariant at this position among the archaeal pols aligned by Edgell et al. (1997), and highly conserved in the pol α family aligned by Braithwaite & Ito (1993).

The model of a ternary complex with dNTP and DNA places residues S489 and R487 in hydrogen-bonding distance from the triphosphate moiety of

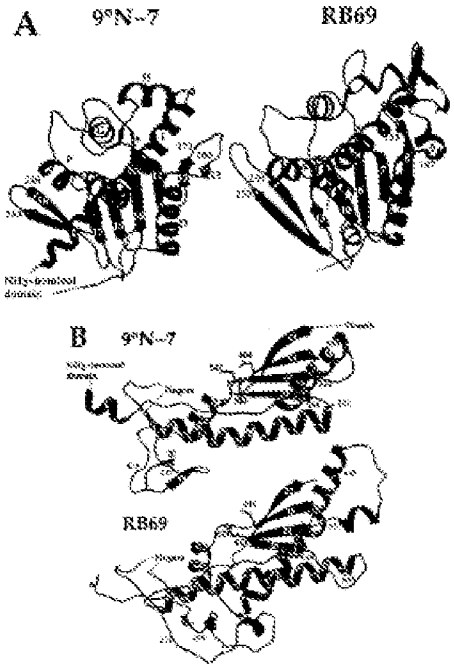


Figure 4. Comparisons of 9°N-7 and RB69 pols to different technologies to indicate loop segments that are shorter in 9°N-7 pol. Least-squares Cα superposition was performed over the region in blue, and the domains were separated for side-by-side comparison. Loop regions are shown in magenta and their relative endpoints are labeled. (a) Comparison of the exonuclease domains. Indicated with purple asterisks are the active site carboxylates (mutated to Ala in the case of the "9°N-less" pol used in this study). (b) Comparison of the palm domains. The three active-site carboxylate groups are depicted with side-chains.

the incoming dNTP. Both of these residues are invariant in the pol α family (Braithwaite & Ito, 1993), and nearly invariant (less exception) among archaeal pols (Edgell *et al.*, 1997). Mutation of the corresponding residues (S488, K490 in Vent (exo-) pol severely decreases enzyme activity (Gardner & Jack, 1999).

Concerted domain movement

The difference in position of the fingers subdomain in 9°N-7 and RB69 pols is part of a larger conformational change involving the 3′-5′ exonuclease and NH_2-terminal domains. Comparing these two pol structures shows that in one of the pols, an essentially rigid-body rotation has occurred involving three of the five (sub)domains. This concerted movement affects both the position of the fingers relative to the pol active site (open versus closed conformation), as well as the position of the exonuclease active site relative to the pol active site. The 9°N-7 and RB69 pol structures may approximate different states along the reaction pathway corresponding to DNA synthesis and 3′-5′ exonucleolytic proofreading activities.

When these two polymerases are aligned in the palm (the blue region in Figure 4b), the exonuclease and fingers are displaced between the proteins (Figure 5(a)). If the enzymes are aligned in the exonuclease domain (see Figure 5(a)), the fingers superimpose almost exactly (Figure 5(b)). Moving from a palm to an exonuclease-based alignment also brings the first module (residues 1-31) of the NH_2-terminal domains into identical positions (not shown). The joint motion of the first NH_2-terminal module and the exonuclease may reflect the need to maintain ionic networks at the interface. There are two five-membered ionic net-

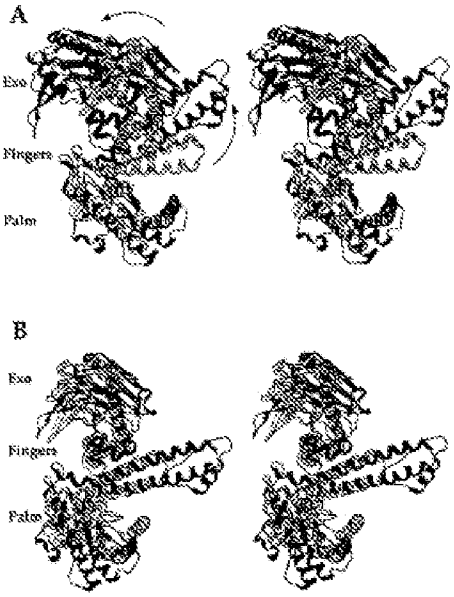


Figure 6. Least-squares Cα superpositions of 9°N-7 and B869 pols in the (a) palm subdomain or (b) exonuclease domain. The 9°N-7 pol backbone is shown in yellow, and its active-site carboxylate groups in gold. The B869 pol backbone is drawn in green, and its active-site residues in magenta. The central β-sheets of the exonuclease domain is light blue (9°N-7 pol) or dark blue (B869 pol) to allow tracking of the domain motion. The precise regions used in the palm and exonuclease superpositions are shown in Figure 4. The NH_2-terminal domain has been omitted for clarity. Arrows in (a) indicate the direction of fingers and exonuclease movement when moving from (a) to (b).

works formed between the first module and exonuclease (Figure 7). In addition, a three-membered network is formed between the third NH_2-module (R346) and the exonuclease (Figure 7). This network is conserved among nearly all archaeal pols (Edgell et al., 1997), but more is present in B869 pol.

Comparison of the Tgo pol structure (Hopfner et al., 1999) with that of 9°N-7 and B869 pols using palm and exonuclease-based superpositions gives results similar to those in Figure 5, providing further support for the notion of a concerted domain movement.

A model was constructed for the B869 pol (Wang et al., 1997) showing how substrate DNA could shuttle between the pol and exonuclease active sites. When 9°N-7 and B869 pols are aligned in the palm, the exonuclease active site in the former is tilted out and away from the pol active site, making it impossible for the DNA to shuttle. The exonuclease position in B886, but not that in 9°N-7 pol, is therefore consistent with an editing conformation. It is interesting that this conformation also means that the fingers are not in position to bind dNTP (see above). Taken together, these considerations suggest that during the replication cycle of family B pols, there is concerted movement of the exonuclease, NH_2-terminal domain, and fingers relative to the catalytic region of the palm.

This concerted movement may be the structural basis for the functional coupling of polymerase and exonuclease domains, which is unique to the pol α family. In this family it is possible to generate site-directed mutations in one domain that exert an indirect, negative effect on the other (Reha-Krantz & Nonay, 1993; Abdus Sattar et al., 1998). This contrasts with pol I pols like KF, where these activities are completely confined to their respective domains (Ollis et al., 1985).

Molecular basis of thermostability

Thermococcus sp. 9°N-7 grows at temperatures of 88–90°C, and its pol has a temperature optimum of 70–80°C (Perler et al., 1996). It has a half-life of 6.7

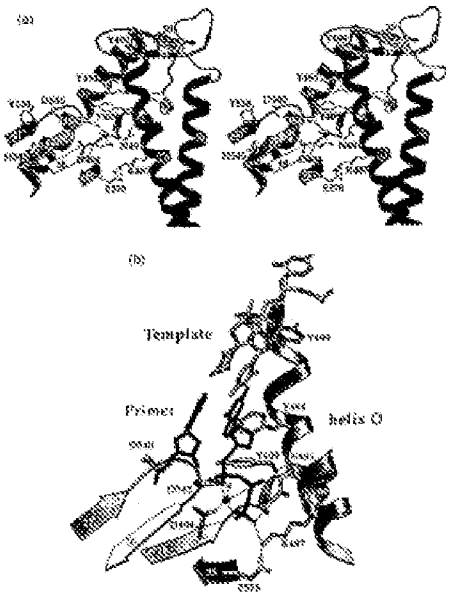


Figure 6. The active site of 9°N-7 pol and a modelled ternary complex. (a) Stereoview of the active site. Residues with indicated side-chains are discussed in the text. Hydrogen bonds are as broken lines, and the two disulfide bridges are shown in white. K487 in this structure is involved in a salt-bridge with E678 of the palm. (b) Model of a ternary complex of 9°N-7 pol. For clarity only the incoming base and the first primer/template base-pair are shown. Hydrogen bonds are shown as broken lines, and metal ions are modelled as green spheres. The 9°N-7 pol and T7 pol ternary complex (Doublié *et al.*, 1998) were superimposed on the palm (0.55 Å rms) for 53 Cα atoms. The mismatch conformation was adopted for D542 and D408 in 9°N-7 pol, and the β turn including D542 was tilted downward, in a motion analogous to that observed between the open enzyme and binary complex structures of *Bacillus stearothermophilus* pol (Kiefer *et al.*, 1998).

homes at 95°C (R.B. Kozyra, unpublished results), whereas *Thermus aquaticus* (Taq) DNA pol has a half-life of 1.6 hours at 95°C (Chien *et al.*, 1976). The structure of 9°N-7 pol indicates a few key strategies for this hyperthermostability, some of which appear general to archaeal DNA pols.

A surprising feature of the 9°N-7 pol is that it contains two disulfide bridges (Figures 3(a) and 6(a)). The potential for the same bridges to form was also observed in Tgo pol (Hopfner *et al.*, 1999). Although not normally the case in Bacteria or Eucarya, an increasing number of cytosolic proteins with disulfide bridges are being discovered in the Archaea (DeDecker *et al.*, 1996; Singleton *et al.*, 1998). The stabilizing role of disulfide bridges has been well documented (Goktoan *et al.*, 1997; Cooper *et al.*, 1982). Introduction of disulfide bridges therefore appears to be a common strategy for archaeal protein stability.

Alignment of a large number of archaeal pols (Edgell *et al.*, 1997) suggests that having at least one of these disulfides is important for their thermostability. In fact, the two-stranded β-sheet containing C442 corresponds to sequence motif D in archaeal pols (Edgell *et al.*, 1997). Based on whether Cys is present in the corresponding positions, all the pols discussed by Edgell *et al.* (1997) are predicted to have at least one of the two disulfide bridges seen in 9°N-7 pol, with the exception of *M. mobile* and *S. shibatae* B3 pols. The thermostability of *M. mobile* pol may be partly caused by a lack of disulfide bridges. The *S. shibatae* B3 pol, like the *S. solfataricus* P2 B5 pol, is highly divergent in sequence from other archaeal pols, and it is unclear whether either of these functions in vivo (Edgell *et al.*, 1997).

An increased number of salt-bridges relative to mesostable homologs is often cited as a determinant of protein thermostability (DeDecker *et al.*, 1996; Kumsforter *et al.*, 1995; Chan *et al.*, 1998; Hennig *et al.*, 1998). The 9°N-7 pol shows a substantial increase in the fraction of charged residues participating in salt-bridges (47 %) compared with RB69 pol (39 %). These results are similar to a thermostability study of *Pyrococcus furiosus* glutamate dehydrogenase (Yip *et al.*, 1995). The authors of

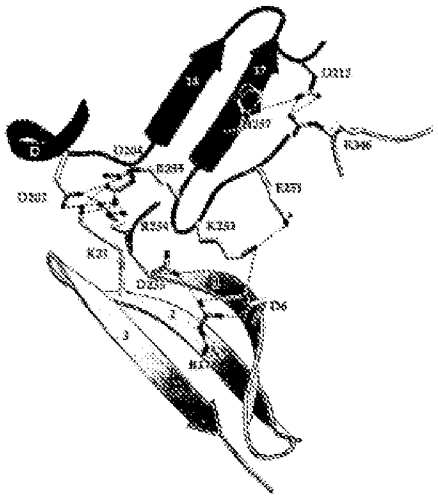


Figure 7. The extensive ionic networks at the interface of the NH_2-terminal and 3′-5′ exonuclease domains.

that study found a marked preference for Arg residues in the ionic interactions of the thermostable enzyme, but no such preference is evident here. The same fraction (46%) of Arg residues is used in ionic interactions in both 9°N-7 and RB69 pols, whereas a much higher proportion of Glu residues participate in salt-bridges in the 9°N-7 pol (83%) compared with RB69 pol (33%).

The number and distribution of salt-bridges within domains does not substantially differ between 9°N-7 and RB69 pols. At the interfaces between (sub)domains, however, the differences in ionic networks are striking. The proportion of ionic interactions at interfaces in the 9°N-7 pol (21%) is over twice that in RB69 pol (9%). The differences lie at the interface of the exonuclease domain with the NH_2-terminal domain (Figure 7), and at the interface of the exonuclease with the thumb, where a two and a three-member ionic network occur in 9°N-7 pol compared with none in RB69 pol (not shown).

Burial of the charged termini of proteins has been cited as another factor that can confer thermostability (Herning *et al.*, 1995). The NH_2-terminal methionine (M1) of 9°N-7 pol is stabilized by a hydrophobic cluster formed by L129, F327, I256, V306, and L341 while the corresponding residue of RB69 pol is completely exposed to solvent. The B-factor for the C^{α} of M1 in 9°N-7 pol is 26 Å^2, whereas for M1 in RB69 pol, it is 83 Å^2. While burial of the N terminus may be important for the thermostability of the 9°N-7 pol, the same does not hold for the C terminus. The last 25 residues are not visible in the electron density, similar to the case of RB69 pol. The solvent accessibility of the C terminus of these pols may reflect the need for this region to interact with a processivity accessory protein, which is known to be the case in the T4 replication complex (Benkovic *et al.*, 1996).

Another common strategy for protein thermostability is to lower the solvent-accessible surface area of the protein and to increase the proportion of buried structure (Kumeharker *et al.*, 1995; Chan *et al.*, 1995). This translates into a more compact structural design. There are at least 13 examples of loop segments in RB69 pol that are much shorter or absent in 9°N-7 pol. Some of the more striking examples are shown in Figure 8. Alignment of 18 archaeal pols (Edgell *et al.*, 1997) indicates that they share practically all of these sequence "deletions". The Pgo pol structure also revealed shortened loop segments relative to RB69 pol (Hopfner *et al.*, 1999). Nevertheless, the overall ratio of solvent-accessible surface area to volume for both 9°N-7 and RB69 pols is the same (8.33). Thus, while lowering the surface area to volume ratio is a common strategy for thermostability, it is not the primary basis for the stability of 9°N-7 pol.

Materials and Methods

Purification, crystallization, and data collection

Thermococcus sp. 9°N-7 polymerase (wild-type and the D141A,E143A exonuclease-deficient mutant) was over-expressed and purified as described (Southworth et al., 1996). Crystallization, cryoprotection, data collection and reduction of native crystals are described (Zhao et al., 1999). Derivatives were prepared by soaking native crystals in stabilization solution (Zhao et al., 1999) supplemented with 22.7 mM sodium ethylmercurithiosalicylate (thimerosal) for 11 days (thimerosal-1), 3.0 mM K_2PtCl_4 for one hour (PtCl-1), 1.5 mM di-μ-iodobis (ethylenediamine)diplatinum (II) nitrate (PIP) for one day (PIP-1), or 1.0 mM Baker's mercurial for 30 hours (BAHg). These crystals were stepped through stabilization solution containing 5% (five minutes), 10% (five minutes), and 20% sucrose (one to five hours). Additional derivatives were collected with the improved cryoprotection procedure reported (Zhao et al., 1999) by soaking native crystals in 23.0 mM thimerosal for 6.3 days (thimerosal-2), 3.0 mM K_2PtCl_4 for seven days (PtCl-2), and 1.5 mM PIP for 38 hours (PIP-2).

Structure determination

The structure of the D141A,E143A mutant of 9°N-7 polymerase was determined by the method of multiple isomorphous replacement (MIR). A number of native and derivative crystals were used to solve the structure because of problems with non-isomorphism (Table 1). Three native datasets were collected from single crystals. NAT-1 was mounted in the liquid nitrogen stream directly from cryoprotectant, whereas NAT-2 and -3 were flash-frozen in liquid nitrogen prior to recording. The crystals belong to space group $P2_12_12_1$ with unit cell dimensions of approximately $a = 96.3$ Å, $b = 110.3$ Å, $c = 112.2$ Å (for NAT-3). One molecule is present per asymmetric unit, giving a solvent content of approximately 60%.

A difference Patterson map of thimerosal-1 was calculated using the program FFT in the CCP4 suite (CCP4, 1994). One heavy-atom site for this derivative was identified with the program RSPS (Knight, 1989). This site was used to calculate initial phases for NAT-1 at 5 Å resolution using the program MLPHARE (Otwinowski, 1991). Difference Fourier synthesis with the initial phases revealed three sites for the PtCl-1 derivative. Two more sites for this derivative were discovered with the phases derived from both thimerosal-1 and PtCl-1. The correct handedness of the phasing information from these derivatives was determined using MLPHARE, and anomalous scattering data from the derivatives were included in the phase calculation. Three sites for the BAHg derivative and four sites for PIP-1 were obtained from difference Fourier maps calculated to 5 Å resolution. All of these heavy-atom sites were included in subsequent phase calculations with NAT-1. The high-resolution phasing limit was extended to 3.5 Å. Because of the high solvent content in the crystals, use of the solvent-flattening program DM (Cowtan, 1994), in combination with histogram matching, improved the phases substantially. A polyalanine model was built into the improved electron density map of NAT-1 with the program O (Jones & Kjeldgaard, 1993) and refined in the program X-PLOR (Brünger, 1992). Phase combination using the program SIGMAA (Read, 1986) further improved the map during building and refinement.

Identification of side-chain densities was possible only after collecting a higher-resolution native dataset (NAT-2), along with diffraction data for three more derivatives obtained under improved cryoprotection conditions (thimerosal-2, one site; PtCl-2, four sites; PIP-2, five sites). These derivatives were used to calculate MIR phases of NAT-2 to 3.0 Å resolution. Partial model phases of NAT-2 were calculated using the refined polyalanine model derived from NAT-1. Because of significant differences in unit cell dimensions between NAT-1 and 2, it was first necessary to subject NAT-2 to rigid-body refinement against NAT-1 in X-PLOR. Combination of the polyalanine model phases and MIR phases with SIGMAA improved the electron density map. Model building, refinement, and phase combination were continued until a complete polyalanine model could be built. In the final stage of refinement, NAT-3 was used to extend the resolution limit to 2.2 Å and water molecules were added.

Coordinate files and illustrations

The *Thermococcus* sp. 9°N-7 polymerase atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank under the accession code 1QHT. The RB69 coordinates used for comparisons in this manuscript are those of the orthorhombic crystal form (accession code 1WAF). Figures were prepared within the IRIS Showcase program (Silicon Graphics, Inc.) entirely (1(b), 2, 3(a) and 3(b)) or with images imported from MOLSCRIPT (1(a)) (Kraulis, 1991) or SETOR (3(c), 4-7) (Evans, 1993).

Acknowledgments

We thank New England Biolabs, Inc. for their collaboration on this project especially S. Kay Williams and Rebecca Kucera for protein purification and Francine Perler and William Jack for helpful discussions. We thank Airlie McCoy, David Owen, Daniela Rhodes, Jean-Luc Jestin, and Hermann Hellinga for critical comments on the manuscript. We also thank Airlie McCoy, Jeff Taylor, and Scott Johnson for assistance in figure preparation. This work was supported by grants to L.S.B. from the ACS (NP-925A), the North Carolina Biotechnology Center, and the Searle Scholars Program.

References

Abdus Sattar, A. K., Lin, T. C., Jones, C. & Konigsberg, W. H. (1996). Functional consequences and exonuclease kinetic parameters of point mutations in bacteriophage T4 DNA polymerase. *Biochemistry*, 35, 16621-16629.

Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily, and monophyly from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA*, 93, 9188-9193.

Beese, L. S. & Steitz, T. A. (1991). Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* 10, 25-33.

Berdis, A. J., Soumillion, P. & Benkovic, S. J. (1996). The carboxyl terminus of the bacteriophage T4 DNA

USSN 10/089,027

Atty. Docket No. 2007651-0001

4567604v1

polymerase is required for holoenzyme complex formation. Proc. Natl Acad. Sci. USA, 95, 12822-12827.
Blanco, L., Bernad, A. & Salas, M. (1992). Evidence favouring the hypothesis of a conserved 3'-5' exonuclease active site in DNA-dependent DNA polymerases. Gene, 112, 139-144.
Braithwaite, D. K. & Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucl. Acids Res. 21, 787-802.
Brünger, A. T. (1992). X-PLOR Version 3.1: A System for X-ray Crystallography and NMR, Yale University Press, New Haven, CT.
Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D. et al. (1996). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science, 273, 1058-1073.
Brautigam, C. A. & Steitz, T. A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrates. Curr. Opin. Struct. Biol. 8, 54-63.
Burd, C. G. & Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. Science, 265, 615-621.
Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. W. & Rees, D. C. (1995). Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. Science, 267, 1463-1469.
Collaborative Computational Project No. 4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallog. sect. D, 50, 760-763.
Cooper, A., Eyles, S. J., Radford, S. E. & Dobson, C. M. (1992). Thermodynamic consequences of the removal of a disulphide from hen lysozyme. J. Mol. Biol. 225, 939-943.
Copeland, W. C., Lam, N. K. & Wang, T. S.-F. (1993). Fidelity studies of the human DNA polymerase α: the most conserved region among α-like DNA polymerases is responsible for metal-induced infidelity in DNA synthesis. J. Biol. Chem. 268, 11041-11049.
Copeland, W. C. & Wang, T. S.-F. (1993). Mutational analysis of the human DNA polymerase α: the most conserved region in α-like DNA polymerases is involved in metal-specific catalysis. J. Biol. Chem. 268, 11028-11040.
Cowtan, K. (1994). "DM": an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM News. Protein Crystallog. 31, 34-38.
DeDecker, B. S., O'Brien, R., Fleming, P. J., Geiger, J. H., Jackson, S. P. & Sigler, P. B. (1996). The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. J. Mol. Biol. 264, 1072-1084.
Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. (1990). An attempt to unify the structure of polymerases. Protein Eng. 3, 461-467.
DeLong, E. F., Wu, K. Y., Prezelin, B. B. & Jovine, R. V. M. (1994). High abundance of Archaea in Antarctic marine picoplankton. Nature, 371, 695-697.
Derbyshire, V., Pinsonneault, J. K. & Joyce, C. M. (1995). Structure-function analysis of 3'-5' exonuclease of DNA polymerases. Methods Enzymol. 262, 363-385.
Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. Nature, 391, 251-258.
Doublié, S., Sawaya, M. R. & Ellenberger, T. (1999). An open and closed case for all polymerases. Structure, 7, R31-R35.
Edgell, D. R. & Doolittle, W. F. (1997). Archaea and the origin(s) of DNA replication proteins. Cell, 89, 995-998.
Edgell, D. R., Klenk, H.-P. & Doolittle, W. F. (1997). Gene duplications in evolution of archaeal family B DNA polymerases. J. Bacteriol. 179, 2632-2640.
Eom, S. H., Wang, J. & Steitz, T. A. (1996). Structure of Taq polymerase with DNA at the polymerase active site. Nature, 382, 278-281.
Evans, S. V. (1993). SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. J. Mol. Graphics, 11, 134-138.
Gandhi, V. & Plunkett, W. (1995). Cytotoxicity, metabolism, and mechanisms of action of 2,2'-difluorodeoxyguanosine in Chinese hamster ovary cells. Cancer Res. 55, 1517-1524.
Gardner, A. F. & Jack, W. E. (1999). Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. Nucl. Acids Res. 27, 2545-2553.
Gokhale, R. S., Agarwalla, S., Francis, V. S., Santi, D. V. & Balaram, P. (1994). Thermal stabilization of thymidylate synthase by engineering two disulfide bridges across the dimer interface. J. Mol. Biol. 235, 89-94.
Hausner, M., Mayer, C., Kohrer, C., Goldener, G., Coolner, P. & Piendl, W. (1994). Autogenous translational regulation of the ribosomal MvaL1 operon in the archaebacterium Methanococcus vannielii. J. Bacteriol. 176, 409-418.
Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J. N. (1995). 2.0 Å structure of indole-3-glycerol phosphate synthase from the hyperthermophile Sulfolobus solfataricus: possible determinants of protein stability. Structure, 3, 1295-1306.
Hopfner, K.-P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R. & Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius. Proc. Natl Acad. Sci. USA, 96, 3600-3605.
Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science, 282, 1669-1675.
Huang, P. & Plunkett, W. (1995). Fludarabine- and gemcitabine-induced apoptosis: incorporation of analogs into DNA is a critical event. Cancer Chemother. Pharmacol. 36, 181-188.
Jones, T. A. & Kjeldgaard, M. (1993). O Version 5.9: The Manual, Uppsala University, Uppsala, Sweden.
Joyce, C. M. & Steitz, T. A. (1994). Function and Structure relationships in DNA polymerases. Annu. Rev. Biochem. 63, 777-822.
Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers, 22, 2577-2637.
Keating, M. J., McCredie, K. B., Bodey, G. P., Smith, T. L., Gehan, E. & Freireich, E. J. (1982). Improved prospects for long-term survival in adults with acute myelogenous leukemia. J. Am. Med. Assoc. 248, 2481-2486.
Kiefer, J. R., Mao, C., Hansen, C. J., Basehore, S. L., Hogrefe, H. H., Braman, J. C. & Beese, L. S. (1997). Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 Å resolution. Structure, 5, 95-108.

Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998). Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature*, 391, 304-307.

Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W. & Steitz, T. A. (1995). Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature*, 376, 612-616.

Kong, H., Kucera, R. B. & Jack, W. E. (1993). Characterization of a DNA polymerase from the hyperthermophile Archaea *Thermococcus litoralis*. *J. Biol. Chem.* 268, 1965-1975.

Korndorfer, I., Steipe, B., Huber, R., Tomschy, A. & Jaenicke, R. (1995). The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. *J. Mol. Biol.* 246, 511-521.

Korolev, S., Nayal, M., Barnes, W. M., Di Cera, E. & Waksman, G. (1995). Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5 Å resolution. *Proc. Natl Acad. Sci. USA*, 92, 9264-9268.

Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* 17, 7514-7525.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature*, 313, 762-766.

Otwinowski, Z. (1991). In *Isomorphous Replacement and Anomalous Scattering* (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds), vol. 80, Science and Engineering Research Council, Warrington, UK.

Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science*, 264, 1930-1935.

Perler, F. B., Kumar, S. & Kong, H. (1996). Thermostable DNA polymerases. *Advan. Protein Chem.* 48, 377-435.

Priestle, J. P. (1988). A program to produce both detailed and schematic drawings for protein structures. *J. Appl. Crystallog.* 21, 946-950.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog. sect. A*, 42, 140-149.

Rees, D. C. & Adams, M. W. W. (1995). Hyperthermophiles: taking the heat and loving it. *Structure*, 3, 251-254.

Reha-Krantz, L. J. & Nonay, R. L. (1993). Genetic and biochemical studies of bacteriophage T4 DNA polymerase 3'-5' exonuclease activity. *J. Biol. Chem.* 268, 27100-27108.

Robertson, L. E. & Plunkett, W. (1993). High-dose cytosine arabinoside in chronic lymphocytic leukemia: a clinical and pharmacologic analysis. *Leuk. Lymphoma*, 10, 43-48.

Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. & Steitz, T. A. (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nature Struct. Biol.* 4, 215-222.

Shimmin, L. C. & Dennis, P. P. (1989). Characterization of the L11, L1, L10, and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium *Halobacterium cutirubrum*. *EMBO J.* 8, 1225-1235.

Singleton, M. R., Isupov, M. N. & Littlechild, J. A. (1999). X-ray structure of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon *Thermococcus litoralis*. *Structure*, 7, 237-244.

Southworth, M. W., Kong, H., Kucera, R. B., Ware, J., Jannasch, H. W. & Perler, F. B. (1996). Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on *Thermococcus* sp. 9°N-7 and mutations affecting 3'-5' exonuclease activity. *Proc. Natl Acad. Sci. USA*, 93, 5281-5285.

Tabor, S. & Richardson, C. C. (1995). A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl Acad. Sci. USA*, 92, 6339-6343.

Tuerk, C., Eddy, S., Parma, D. & Gold, L. (1990). Autogenous translational operator recognized by bacteriophage T4 DNA polymerase. *J. Mol. Biol.* 213, 749-761.

Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H. & Steitz, T. A. (1996). Crystal structures of an NH_2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. *Biochemistry*, 35, 8110-8119.

Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997). Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. *Cell*, 89, 1087-1099.

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA*, 87, 4576-4579.

Wong, S. W., Wahl, A. F., Yuan, P. M., Arai, N., Pearson, B. E., Arai, K., Korn, D., Hunkapiller, M. W. & Wang, T. S.-F. (1988). Human DNA polymerase α gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. *EMBO J.* 7, 37-47.

Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. & Rice, D. W. (1995). The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure*, 3, 1147-1158.

Zhao, M., Mao, C., Rodriguez, A. C., Kiefer, J. R., Kucera, R. B. & Beese, L. S. (1998). Crystallization and preliminary diffraction analysis of a hyperthermostable DNA polymerase from a *Thermococcus* archaeon. *Acta Crystallog. sect. D*, 54, 998-999.

Edited by D. Rees

(Received 15 June 1999; received in revised form 24 March 2000; accepted 24 March 2000)

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 278, No. 12, Issue of March 21, pp. 11004–11013, 2003
Printed in U.S.A.

Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase*

Received for publication, July 29, 2002, and in revised form, December 23, 2002
Published, JBC Papers in Press, December 26, 2002, DOI 10.1074/jbc.M208288200

Andrew F. Gardner‡, Catherine M. Joyce§, and William E. Jack‡¶

From ‡New England Biolabs Inc., Beverly, Massachusetts 01915 and the §Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520

Comparative kinetic and structural analyses of a variety of polymerases have revealed both common and divergent elements of nucleotide discrimination. Although the parameters for dNTP incorporation by the hyperthermophilic archaeal Family B Vent DNA polymerase are similar to those previously derived for Family A and B DNA polymerases, parameters for analog incorporation reveal alternative strategies for discrimination by this enzyme. Discrimination against ribonucleotides was characterized by a decrease in the affinity of NTP binding and a lower rate of phosphoryl transfer, whereas discrimination against ddNTPs was almost exclusively due to a slower rate of phosphodiester bond formation. Unlike Family A DNA polymerases, incorporation of 3′-(2-hydroxyethoxy)methyl-dNTP triphosphates (where N is adenine, cytosine, guanine, or thymine; acyclo-NTPs) by Vent DNA polymerase was enhanced over ddNTPs via a 30-fold increase in phosphoryl transfer rate. Furthermore, a mutant with increased propensity for nucleotide analog incorporation (Vent^{A488E} DNA polymerase) had unaltered dNTP incorporation while displaying enhanced nucleotide analog binding affinity and rates of phosphoryl transfer. Based on kinetic data and available structural information from other DNA polymerases, we propose active site models for dNTP, ddNTP, and acyclo-NTP selection by hyperthermophilic archaeal DNA polymerases to rationalize structural and functional differences between polymerases.

All free living organisms encode several DNA polymerases that are jointly responsible for the replication and maintenance of their genomes. Thereby ensuring accurate transmission of genetic information (1–3). The majority of identified DNA polymerases can be classified into Families A, B, C, and Y according to amino acid sequence similarities to *Escherichia coli* polymerases I, II, III, and DNA, respectively (4, 5). Additional families have been identified, including the two-subunit replicative DNA polymerases from hyperthermophilic Archaea (Family D) (6) and eukaryotic DNA polymerase β and terminal transferases (Family X) (6).

Structural and kinetic analyses of Family A (7–14) and Family B (15–25) DNA polymerases have increased our understanding of nucleotide selection and incorporation mechanisms. Although amino acid sequences diverge between these two families, the structures of Family A and B DNA polymerases share recognizable fingers, thumb, and palm subdomains that allow comparisons of structural elements important for function (8, 26). In the case of Family A DNA polymerases from bacteriophage T7, *Escherichia coli* (Klenow fragment, large fragment of DNA polymerase I), and *Thermus aquaticus*, as well as the Family B DNA polymerases from bacteriophage RB69, incorporation of the structural information is complemented by steady-state and pre-steady-state kinetic studies, allowing a detailed description of the polymerization pathway. Reaction parameters describing the discrimination against naturally occurring nucleotide analogs encountered in vivo, such as NTPs, or structural nucleotide analogs, such as ddNTPs and dye-labeled ddNTPs (13, 25–30), have added insights into the basis for nucleotide discrimination.

Hyperthermophilic archaeal DNA polymerases have not been scrutinized in such detail, hampering a complete characterization and comparison with other polymerases. Family B DNA polymerases from hyperthermophilic Archaea *Thermococcus* sp. 9°N (22), *Thermococcus gorgonarius* (18), and *Pyrococcus kodakaraensis* KOD1 (24) and bacteriophage RB69 (23) have high sequence and structural homologies and provide a framework for analysis of active site structure and function in this enzyme family (Fig. 1). Furthermore, steady-state kinetic studies have identified hyperthermophilic DNA polymerase residues important for polymerization and exonuclease activities and for nucleotide binding (18, 29, 31–36). Nucleotide analogs have also been important in identifying dNTP recognition determinants important to the polymerase reaction (32–38) and have proven useful in a variety of molecular biology applications, such as DNA sequencing and detection of single nucleotide polymorphisms (37–41). One group of analogs, 3′-(2 hydroxyethoxy)methyl-N triphosphates (where N is adenine, cytosine, guanine, or thymine; acyNTPs),[1] is particularly intriguing due to the wide spectrum of incorporation efficiency noted to different DNA polymerases, even within the same family of polymerase. For example, within Family B, the herpes simplex virus type 2 and human cytomegalovirus DNA polymerases incorporate acyNTPs more efficiently than dcNTPs, whereas human polymerase α more readily accepts dcNTPs over acyNTPs (42). Such differences have been exploited in drug therapies where infective agents encode polymerases that more readily incorporate acyNTPs than those the host DNA polymerases (43). Hyperthermophilic archaeal DNA poly-

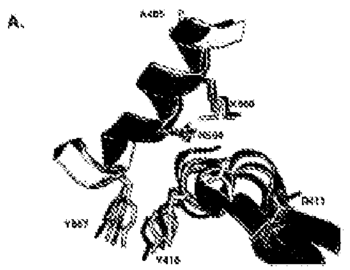
* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Supported by National Institutes of Health Grant GM-22250.

¶ To whom correspondence should be addressed: New England Biolabs Inc., 32 Tozer Rd., Beverly, MA 01915. Tel.: 978-927-5054; Fax: 978-355-2380; E-mail: jack@neb.com.

[1] The abbreviations used are: acyNTP, 3′-(2-hydroxyethoxy)methyl-NTP triphosphate, where N is adenine, cytosine, guanine, or thymine; acyNN, 3′-(2-hydroxyethoxy)methyl-N triphosphate; dCTPαS, 2′-deoxycytidine-5′-O-(1-thiotriphosphate); ddNTPαS, 2′,3′-dideoxyguanine 5′-O-(1-thiotriphosphate); 40DA, 5-carboxy-2,7-dichlorofluorescein; FAM, 6-carboxyfluorescein.

A.



B.

	Region II	Region III
RB69	411 DLTSLYPSII 420	557 INRKLLINSLY 567
Vent	407 DFRSLYPSII 416	487 RAIKLLANSYY 497
9°N	404 DFRSLYPSII 413	485 RAIKILANSFY 495
KOD	404 DFRSLYPSII 413	484 RAIKILANSYY 495
TGO	404 DFRSLYPSII 413	484 RAIKILANSFY 495

FIG. 1. Alignments of Family B DNA polymerase active sites. [illegible]

[illegible]

EXPERIMENTAL PROCEDURES

[illegible]

USSN 10/089,027

Atty. Docket No. 2007651-0001

4567604v1

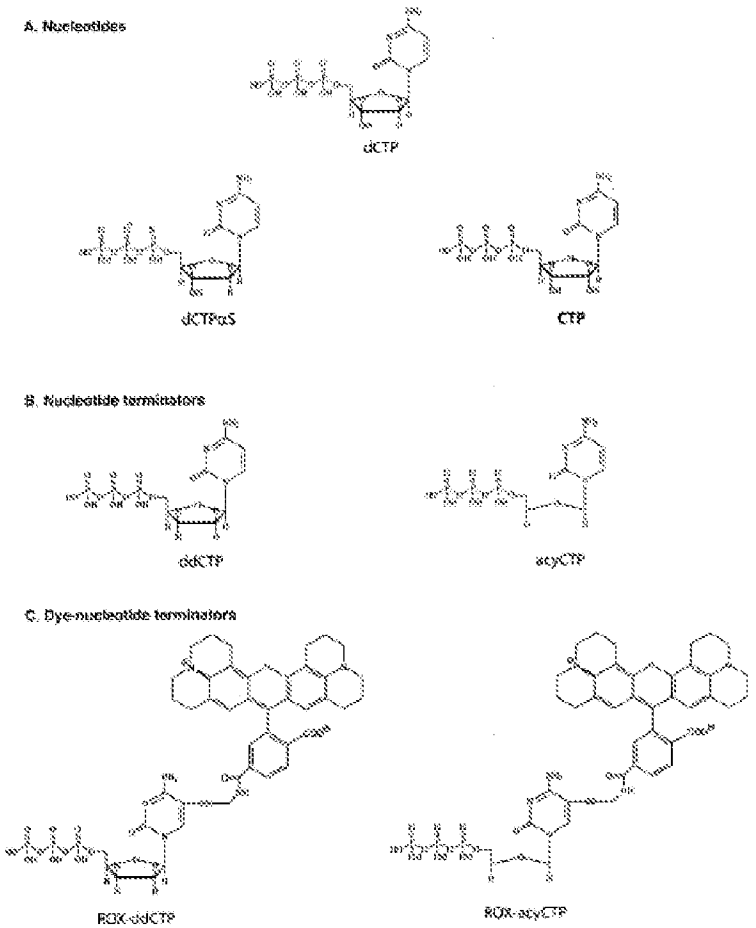


FIG. 2. Nucleotide and nucleotide analogs used the study of Vent DNA polymerases pre-steady-state kinetic reactions. The maximum rate of nucleotide addition and the dissociation constant for binding were determined with the following nucleotides and nucleotide analogs: dCTP, dCTPαS, and CTP (A), nucleotide terminators ddCTP and acyCTP (B), and dye-labeled nucleotide terminators ROX-ddCTP and ROX-acyCTP (C).

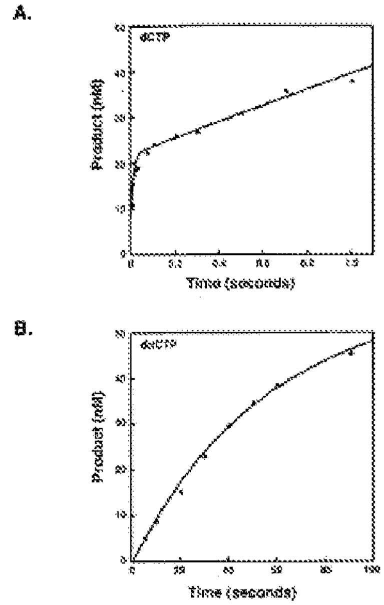
nucleotide or analog concentration, and fitted to the hyperbolic equation $k_{obs} = (k_{pol}[nucleotide])/(K_d + [nucleotide])$, yielding k_{pol}, the maximum rate of nucleotide addition, and K_d, the dissociation constant for nucleotide binding (16). The correction energy difference between dCTP and nucleotide analog incorporation was calculated by Equation 2 (32).

[illegible]

[illegible]

RESULTS

[illegible]



[illegible]

[illegible]

USSN 10/089,027

Atty. Docket No. 2007651-0001

4567604v1

Table 1

Pre-steady-state burst kinetics

The kinetic parameters for Vent and Vent^{A488L} DNA polymerases are from at least two independent determinations and are reported as the average ± S.D. NB, not determined.

Enzyme	$k_{polymerase}$	$K_{d(app)}$	$k_{exonuclease}$	$k_{exonuclease}$
	s^{-1}	μM	s^{-1}	s^{-1}
Vent	30 ± 10	0.20 ± 0.09	0.47 ± 0.05	0.042 ± 0.025
Vent^{A488L}	45 ± 5	0.16 ± 0.03	0.33 ± 0.02	0.038 ± 0.013
KOD[a]	300 ± 40	3.3 ± 0.3	ND	ND
AmpliTaq-CS[b]	80 ± 7	2.5 ± 0.5	ND	ND

[a] Ref. 28.
[b] Ref. 18.

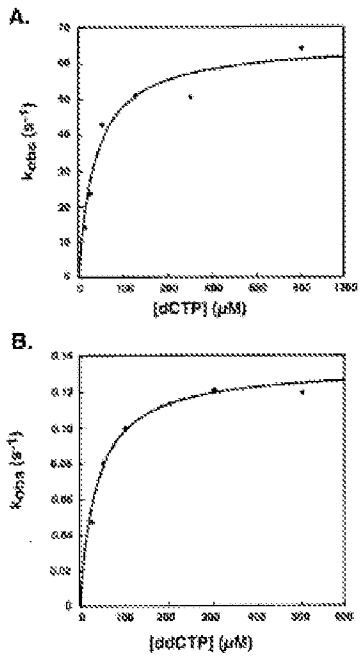


FIG. 6. **Vent DNA polymerase pre-steady-state kinetics of nucleotide and nucleotide analog incorporation.** The dependence of the observed rate (k_{obs}) on nucleotide or nucleotide analog concentration was fit to a hyperbola according to the Michaelis-Menten equation: $k_{obs} = (k_{polymerase}[nucleotide])/(K_d + [nucleotide])$, where k_{obs} is the observed reaction rate, $k_{polymerase}$ is the maximum rate of phosphodiester bond formation, and K_d is the apparent dissociation constant, as described under "Exonuclease Processing." *A*, a fit of the data for dCTP incorporation gave $k_{polymerase} = 75$ μM and $k_{pol} = 65$ s^{-1}. *B*, a fit of the data for ddCTP single turnover gave $K_{d(ddCTP)} = 37$ μM and $k_{pol} = 0.38$ s^{-1}.

dosed by >1500-fold compared with dCTP (Table III).

ROX-ddCTP and ROX-acyCTP Incorporation—Previous studies found ROX-derivatized ddCTP and acyCTP to be more efficient terminators than their unmodified forms when using Vent DNA polymerase (32). Pre-steady-state kinetics revealed higher binding affinities, but slower incorporation kinetics for the ROX derivatives (Table IIb), resulting in only marginal alterations in incorporation selectivity.

Analysis of Enhanced Nucleotide Analog Incorporation by Vent^{A488L} DNA Polymerase—We previously reported enhanced incorporation of nucleotide analogs by Vent^{A488L} DNA polymerase (32, 33). In a burst kinetics experiment, the A488L mutant enzyme gave an initial burst of dCTP incorporation at a rate similar to that seen with the wild-type enzyme ($k_{polymerase} \approx 45$ s^{-1}; >75% active) (Table I). Moreover, the single turnover kinetic parameters for dCTP addition ($K_d \approx 37$ μM and $k_{pol} \approx 58$ s^{-1}) were similar to values derived for the wild-type enzyme (Table II). However, following the initial turnover, the steady-state rate of the A488L mutant polymerase was 8-fold slower than that of the wild-type enzyme ($k_{ss} \approx 5.9$ s^{-1}) (Table I), accounting for the lower specific activity of Vent^{A488L} DNA polymerase (32). As with wild-type Vent DNA polymerase, replacement of dCTP with dCTPαS had little effect on K_d or k_{pol} (Table II). Vent^{A488L} DNA polymerase was less active in pyrophosphorolysis compared with the wild-type enzyme, the result of a 3-fold reduction in PP$_i$ binding affinity and a 3-fold decrease in k_{pyro} (Table IV). Incorporation of TTP, ddCTP, acyCTP, and ROX-ddCTP by Vent^{A488L} DNA polymerase was more efficient (by 8–10-fold) compared with incorporation by the wild-type enzyme; in each case, this is attributable to both increased binding affinity (lower K_d) and faster reaction rates (k_{pol}) (Table II). Incorporation of ROX-acyCTP was largely unaffected by the A488L mutation (Table II).

DISCUSSION

The fluorescent information derived for Vent DNA polymerase incorporation of dCTP confirms and expands earlier steady-state data (31) and places Vent DNA polymerase in the context of other Family A and B DNA polymerases. As with these other polymerases, the steady-state rate for single nucleotide addition is limited by a slow step after phosphodiester bond formation. Previous steady-state measurements using an assay in which all four dNTPs were present gave a k_{cat} value of 10 s^{-1}.[3] This value is higher than the steady-state rate derived here in burst experiments (0.5 s^{-1}), most likely reflecting the higher temperature (72 °C) and, more importantly, the processive synthesis allowed in the earlier studies. In contrast, the experimental design reported here forces the DNA polymerases to act in a distributive manner, *i.e.* dissociating from the DNA before binding another primer-template and incorporating another nucleotide.

The single turnover parameters for Vent DNA polymerase with the normal dCTP substrate are similar to those of other Family A and B polymerases, both exonuclease and thermophilic. As shown in Table III, K_d and k_{pol} values differ by

[3] H. Kong, H. M. S. B. Ekeoma, and W. E. Jack, unpublished data.

TABLE II
Pre-steady-state kinetic constants for nucleotide and nucleotide analog incorporation by Vent and Vent[exo-] DNA polymerases

In almost all cases, the kinetic parameters for Vent and Vent[exo-] DNA polymerases are from at least two independent determinations except where indicated by footnote b and are reported as the mean ± S.D.

Nucleotide	Vent DNA polymerase				Vent[exo-] DNA polymerase			
	K_D	k_{pol}	k_{pol}/K_D	Selectivity[a]	K_D	k_{pol}	k_{pol}/K_D	Selectivity[a]
	μM	s^{-1}	$M^{-1}s^{-1}$		μM	s^{-1}	$M^{-1}s^{-1}$	
dCTP	36 ± 7	66 ± 5	1.8×10^6		72 ± 9	50 ± 5	7.0×10^5	
dCTPαS	320 ± 46	82 ± 13	2.6×10^5	7.4	88 ± 32	30.0 ± 0.9	4.1×10^5	1.9
CTP	3100 ± 500	0.180 ± 0.003	5.8×10^1	3000	390 ± 80	0.70 ± 0.08	1.8×10^3	450
ddCTP	48 ± 7	0.18 ± 0.01	3.8×10^3	490	15 ± 4	0.03 ± 0.02	2.0×10^3	40
araCTP	85 ± 20	7.8 ± 1.1	9.2×10^4	20	24.6 ± 0.4	20 ± 8	8.4×10^5	1.4
[illegible]	30 ± 5	0.059 ± 0.008	2.0×10^3	920	80[b]	0.2[b]	3.0×10^3	30
[illegible]	8.0 ± 0.8	2.00 ± 0.01	2.5×10^5	7	6[b]	1[b]	1.8×10^5	4.5

[a] Selectivity between incorporation of dCTP and other NTPs is the ratio of the efficiency of dCTP incorporation (k_{pol}/K_D) to the efficiency of CTP, ddCTP, or araCTP incorporation.
[b] The kinetic parameters for Vent[exo-] DNA polymerase are from single determinations.

TABLE III
Pre-steady-state kinetic constants for nucleotide analog incorporation by DNA polymerases

The kinetic parameters for Vent and Vent[exo-] DNA polymerases are from at least two independent determinations and are reported as the mean ± S.D. ND, not determined.

Enzyme	dCTP		CTP			ddCTP			araCTP		
	K_D	k_{pol}	K_D	k_{pol}	Selectivity[a]	K_D	k_{pol}	Selectivity[a]	K_D	k_{pol}	Selectivity[a]
	μM	s^{-1}	μM	s^{-1}		μM	s^{-1}		μM	s^{-1}	
Vent	36 ± 7	66 ± 5	3100 ± 500	0.180 ± 0.003	3000	48 ± 7	0.18 ± 0.01	490	85 ± 20	7.8 ± 1.1	20
Vent[exo-]	72 ± 9	50 ± 5	390 ± 80	0.70 ± 0.08	450	15 ± 4	0.03 ± 0.02	40	24.6 ± 0.4	20 ± 8	1.4
RB69	80 ± 18[b]	200 ± 10[b]	16,000 ± 4000[b]	0.36 ± 0.2[b]	64,000[b]	4300 ± 800[b]	0.17 ± 0.02[b]	72,000[b]	ND	ND	ND
Klenow	5.0 ± 2.3[c]	75 ± 12[c]	25 ± 7[c]	0.057 ± 0.005[c]	5800[c]	8.4 ± 4[c]	0.105 ± 0.010[c]	4000[c]	200 ± 50	0.108 ± 0.008	28,000
KlenTaq	55 ± 8[d]	21 ± 4[d]	ND	ND	ND	35 ± 10[d]	0.09 ± 0.002[d]	2200[d]	ND	ND	ND

[a] Selectivity between incorporation of dCTP and other NTPs is the ratio of the efficiency of dCTP incorporation (k_{pol}/K_D) to the efficiency of CTP, ddCTP, or araCTP incorporation.
[b] Ref. 30.
[c] Ref. 26.
[d] Ref. 27.
[e] Ref. 18.

TABLE IV
Pyrophosphorolysis

The kinetic parameters for Vent and Vent[exo-] DNA polymerases are from at least two independent determinations and are reported as the mean ± S.D.

Enzyme	$K_{D(PPi)}$	k_{pyro}	$k_{pyro}/K_{D(PPi)}$	pol/pyro[a]
	μM	s^{-1}	$M^{-1}s^{-1}$	
Vent	340 ± 200	1.10 ± 0.37	3.2×10^3	300
Vent[exo-]	1200 ± 400	0.63 ± 0.21	5.7×10^2	1200
RB69[b]	20,000	0.33	1.5×10^1	200,000
Klenow[c]	200	0.25	1.3×10^3	6000

[a] A comparison of DNA polymerase activity with pyrophosphorolysis (pol/pyro) is given by the ratio of DNA polymerization efficiency ($k_{pol}/K_{D(dNTP)}$) divided by the efficiency of pyrophosphorolysis ($k_{pyro}/K_{D(PPi)}$).
[b] Ref. 30.
[c] Ref. 8.

<10-fold for all polymerases tested, with no clear division between Family A and B DNA polymerases. Furthermore, the Family A Klenow fragment and Family B Vent and RB69 DNA polymerases carry out the reverse reaction of DNA polymerization, pyrophosphorolysis, with similar rates (k_{pyro}), and Klenow fragment and Vent DNA polymerases share comparable PP_i binding constants ($K_{D(PPi)}$) (Table IV). Similarities in nucleotide incorporation kinetics and active site structure underscore the evolution of DNA polymerases to efficiently carry out DNA replication and repair. Significant kinetic differences between the polymerases become apparent only when examining nucleotide analog incorporation and elemental effects, as detailed below.

Ribonucleotides—Despite a similar level of selectivity against NTPs, this discrimination is supplied almost exclusively by elements occurred by k_{pol} for Klenow fragment DNA polymerase, whereas Vent DNA polymerase shows not only k_{pol} effects, but also a 10-fold weaker ground state binding of the nucleotide. RB69 DNA polymerase also shows effects in both K_D and k_{pol}, achieving an even higher discrimination by virtue of a 200-fold weaker ground state binding. Discrimination against NTPs has, in large part, been attributed to a steric clash between the 2′-OH and a conserved side chain in the polymerase active site (28, 30, 33). The kinetic parameters suggest that the steric clash is first encountered in the ground state nucleotide binding by Vent and RB69 DNA polymerases, but does not affect Klenow fragment DNA polymerase until the transition state of the reaction. This could occur, for example, if the K_D term for Klenow fragment DNA polymerase primarily measures binding prior to a conformational shift that engages the 2′-OH sensing machinery.

Dideoxynucleotides—When incorporating ddCTP, RB69 DNA polymerase discriminates at the level of both K_D and k_{pol}. Discrimination by Vent, Klenow fragment, and KlenTaq (truncated Taq DNA polymerase with a 278-amino acid N-terminal deletion (32)) DNA polymerases is almost exclusively in the

steps measured by k_{pol} and not those measured in K_D, with Vent DNA polymerase showing less discrimination than the other two polymerases. This parallel behavior appears to reflect a neutral lack of 3′-OH involvement in ground state substrate binding rather than a measured set of nucleotide contacts.

On the surface, the similarity in k_{pol} values for dNTP incorporation by Vent and Klenow fragment DNA polymerases (30) suggests similar discriminatory mechanisms for these two enzymes, a conclusion reinforced by the absence of an isotopical effect with dNTPαs using either enzyme. The simplest interpretation is the lack of an isotopical effect with α-thio-substituted dNTPs with Klenow fragment and Vent DNA polymerases is that a non-chemical step(s) preceding phosphodiester bond formation is rate-limiting. Similarly, the lack of a significant isotopical effect for Klenow fragment DNA polymerase incorporation of dsNTPαs (32) argues that steps preceding phosphodiester bond formation continue to be rate-limiting for that enzyme. In contrast, the isotopical effect noted for Vent and Vent$^{(exo-)}$ DNA polymerase incorporation of dsNTPαs is an indication that the chemistry of phosphodiester bond formation significantly influences the rate-limiting step for these polymerases.

The k_{pol} rates with both polymerases were significantly slower for dsNTPs than for dNTPs: 5000- and 400-fold for Klenow fragment and Vent DNA polymerases, respectively. In the case of Vent DNA polymerase, this must reflect at least a slowing of the chemical rate, whereas for Klenow fragment DNA polymerase, at least the rate of pre-chemical step(s) must be slowed. Thus, the pre-chemical rate for Vent DNA polymerase dsNTP incorporation is at least 10-fold faster than comparable steps for Klenow fragment DNA polymerase.

Conserved amino acids positioned within either Family A or B DNA polymerase active sites probe for correctly base-paired substrates and coordinately align phosphates into a geometry required for phosphoryl transfer. As observed by Franklin et al. (33) and Yang et al. (28) in the RB69 DNA polymerase ternary ground structure (and by analogy, in the Vent DNA polymerase active site) (Fig. 1), the dNTP discriminator moiety assumes a favorable 3′-endo sugar conformation. This conformation is maintained by hydrogen bonds between the 3′-OH and a main chain amide (corresponding to Vent DNA polymerase position 412) and a non-bridging β-phosphate oxygen (Fig. 5, A and B). Nucleoside α-, β-, and γ-phosphates are further stabilized by direct or water-mediated hydrogen bonds with active site residues (Fig. 5, A and B). The absence of the 3′-OH on dsNTPs disrupts hydrogen bonding with the β-phosphate (and main chain amide), potentially increasing the activation energy required to orient the α-phosphorus for phosphoryl transfer (Fig. 5C). Indeed, the measured energetic difference between dNTP and dsNTP incorporation (12 kJ mol^{-1}) is equivalent to that expected for the loss of at least two hydrogen bonds to the dsNTP transition state (34).

Although the active site binding network differs in the Family A Klenow fragment DNA polymerase the dNTP 3′-OH contributes 21 kJ mol^{-1} to transition state stabilization, accounting for inefficient dsNTP incorporation (27). This energy loss is counteracted in the closely related T7 DNA polymerase active site by a hydroxyl group at Tyr^{526} (Klenow fragment DNA polymerase has Phe in the analogous position) that can facilitate a hydrogen bond to stabilize the dsNTP β-phosphate in the transition state, re-establishing a hydrogen bonding network similar to interactions formed by dNTP (12). As a result, T7 DNA polymerase selectivity between dNTP and dsNTP is greatly reduced, as is the selectivity of the analogous Phe-to-Tyr evolution in both Klenow fragment and *Taq* DNA polymerases (35).

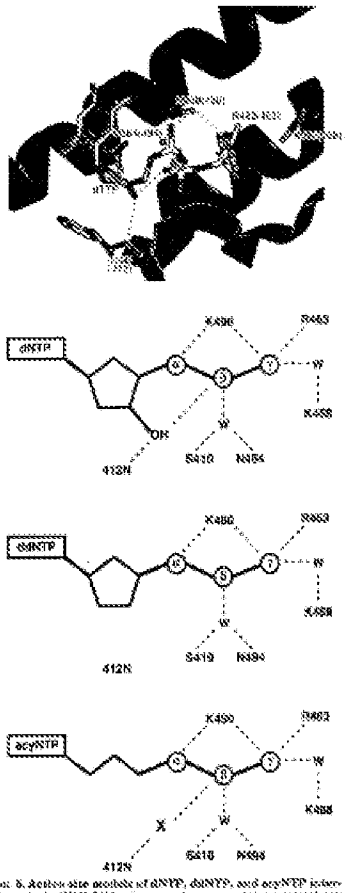


FIG. 5. Active site models of dNTP, dsNTP, and acyNTP interactions. *A*, the RB69 DNA polymerase ternary complex structure shows active site interactions that stabilize the substrate dNTP (33). Vent DNA polymerase numbering is shown in parentheses. *B*, a schematic of the Vent DNA polymerase active site interactions that stabilize the dNTP transition state is presented in a two-dimensional view. W, water molecule. *C*, a model of the Vent DNA polymerase active site with dsNTP bound suggests loss of hydrogen bonding with the Tyr^{412} main chain amide, dNTP 3′-OH, and non-bridging β-phosphate. *D*, a model for the binding of acyNTP suggests that, in the absence of a sulfur ring, a noncatalytic hydrogen bond between the Tyr^{412} main chain amide and non-bridging β-phosphate.

Acyclonucleotides—Divergence between polymerases is also noted upon acyCTP addition, again suggesting divergent mechanisms for nucleotide recognition and incorporation between the polymerases. Similar to ddNTPs, acyNTPs lack the 3′-OH required to establish a hydrogen bonding network between the main chain amide of Tyr[412] and β-phosphate of the substrate (Fig. 5*B*). Klenow fragment DNA polymerase displays a strong discrimination in both K_d and k_{pol} terms, resulting in a selectivity value of 32,880. In this case, the efficiency of acyNTP incorporation is nearly as low as for an incorrect base pair (k_{pol}/K_d = 240 and 360 $M^{-1}s^{-1}$, respectively) (Table III) (60). A strong bias against acyNTP incorporation has also been noted for *Taq* DNA polymerase (33).

In contrast, acyNTPs are incorporated by hyperthermophilic archaeal DNA polymerases with only 10-fold lower efficiency than dNTPs. By analogy with ddNTP incorporation by T7 DNA polymerase, it seems reasonable that the space normally occupied by the sugar 2′- and 3′-carbons and associated substituents would be accessible to water molecules, metals, or protein side chains that might establish interactions to compensate for those disrupted by the missing 3′-OH group. The difference in activation free energy between ddNTP and acyNTP incorporation ($\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}_{ddNTP} - \Delta G^{\ddagger}_{acyNTP}$ = 10 kJ mol^{-1}) is equivalent to a gain of two additional hydrogen bonds, which could be provided by hydrogen bonding between the main chain 412 amide, a putative water, and an acyNTP β-phosphate non-bridging oxygen to mimic interactions that exist in the dNTP active site (Fig. 5*B*). At the same time, we cannot rule out stabilizing interactions arising from residues near the active site normally excluded by the ribose 2′- and 3′-carbons that are absent in acyNTP. Clearly, three-dimensional structural analysis will be necessary for a full understanding of the interactions important for Vent DNA polymerase incorporation of acyNTPs.

Dye-substituted Nucleotides—Dye-substituted nucleotides have been useful in a variety of analytical applications (40, 43). Not surprisingly, given the diversity of dye structures and charges, dye-substituted nucleotides are accepted by DNA polymerases with varying efficiencies (28, 29, 33, 36). Previous studies identified nucleotide derivatives containing the fluorescent dye ROX as being more efficiently incorporated by Vent DNA polymerase than the parental nucleotides lacking the dye (28, 35). In the current kinetic studies, the magnitude of enhanced dye-substituted terminator addition was much lower than previously estimated in semiquantitative gel titration assays, even though those same gel titration assays give good agreement with the relative incorporation efficiency of ddNTP and acyNTP substrates (33). ROX substitution of the nucleotide results in a 5–15-fold lower K_d, suggesting that contacts in or adjacent to the Vent DNA polymerase active site stabilize dye binding. However, at the same time, k_{pol} is reduced, suggesting that one consequence of the enhanced binding is to slow nucleotide addition. Thus, substrate incorporation is a balance of both binding and catalysis: a substrate bound with too high affinity requires higher activation energy for efficient turnover by the polymerase.

Vent[A488L] DNA Polymerase Pre-steady-state Kinetics—Previous studies identified a Vent DNA polymerase variant (A488L) with enhanced nucleotide analog incorporation properties (28, 33). Correct dNTP incorporation by Vent[A488L] DNA polymerase is characterized by similar binding affinity (K_d), nucleotide transfer rate (k_{pol}), and rate-limiting step compared with Vent DNA polymerase, presumably reflecting the conservation of residues actively involved in coordinating the incoming dNTP. In contrast, each of the nucleotide analogs tested with Vent[A488L] DNA polymerase have higher binding affinity and faster rates of phosphoryl transfer than the unmodified polymerase. Energy differences between Vent and Vent[A488L] DNA polymerase incorporation of ddNTP or acyNTP are modest ($\Delta\Delta G^{\ddagger}_{ddNTP}$ = 4.5 kJ mol^{-1} and $\Delta\Delta G^{\ddagger}_{acyNTP}$ = 4.8 kJ mol^{-1}), suggesting that subtle hydrophobic or hydrogen bond-mediated effects could account for enhanced analog incorporation.

One hypothesis to account for these effects envisions the A488L variant as lying closer to the activated conformation, thus facilitating incorporation of analogs. The residue analogous to Ala[488] in the RB69 DNA polymerase crystal structure points away from the active site and lies at the interface between the solid core of the polymerase and an α-helix that must make a 60° rotation to form the closed complex (Fig. 1). In Vent DNA polymerase, positioning a larger leucine residue at the position normally occupied by alanine in the α-helix may shift equilibrium from the open toward the closed conformation, thus reducing the activation energy for both binding and nucleotide transfer. This comes at a price: burst experiments demonstrate that subsequent turnover by the A488L variant is inhibited, perhaps reflecting hindrance of the transition from closed to open states required for release and/or binding of the template and dNTP. This proposal does not, however, easily account for the fact that pre-steady-state kinetics for the natural substrates are unaltered in this variant. Alternatively, resolution of this discrepancy may lie in the greater ability of this variant to overcome distortions in the nucleotide-binding site, distortions that are not present when the natural nucleotide is bound.

In summary, from these comparative studies, we observed that kinetics of dNTP incorporation pathways are conserved among Family A and B DNA polymerases despite diversity in primary amino acid sequence, thermostability, fidelity, and biological roles. However, differences in acyNTP and other nucleotide analog catalytic efficiencies in Klenow fragment, Vent, and other DNA polymerases illuminate fundamental differences underlying the kinetic pathway for DNA polymerization. As more DNA polymerases are studied kinetically, it is apparent that subtle structural variations in the active site influence how nucleotides are bound and positioned for catalysis.

Acknowledgments—We are grateful to Charles Richardson for overseeing this research as part of the Master of Liberal Arts Program at Harvard University (to A. F. G.). We also thank Phil Buzby for providing dye-labeled compounds and for helpful discussions, John Buzzini (Applied Biosystems) for useful discussions and technical advice, Nicole Nichols, Sriharsha Pradhan, and Tom Evans for critical review of the manuscript, and Chris Boucher, Lucia Greenough, and Julie Moore for providing expert technical assistance. We are also indebted to Don Comb for fostering a supportive research environment at New England Biolabs Inc.

REFERENCES

1. Kornberg, A. (1980) *DNA Replication*, pp. 85–95, W. H. Freeman & Co., San Francisco
2. Joyce, C. M., and Steitz, T. A. (1994) *Annu. Rev. Biochem.* 63, 777–822
3. Steitz, T. A. (1999) *J. Biol. Chem.* 274, 17395–17398
4. Brautigam, C. A., and Steitz, T. A. (1998) *Curr. Opin. Struct. Biol.* 8, [illegible]
5. Filée, J., Forterre, P., Sen-Lin, T., and Laurent, J. (2002) *J. Mol. Evol.* 54, 763–773
6. Ghosh, [illegible], Ghosh, K., Tsai, R., Kucera, R., and [illegible] (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, [illegible]
7. Polesky, A. H., Steitz, T. A., Grindley, N. D. F., and Joyce, C. M. (1990) *J. Biol. Chem.* 265, 14579–14591
8. Patel, S. S., Wong, I., and Johnson, K. A. (1991) *Biochemistry* 30, 511–525
9. Dahlberg, M. E., and Benkovic, S. J. (1991) *Biochemistry* 30, 4835–4843
10. Polesky, A. H., Dahlberg, M. E., Benkovic, S. J., Grindley, N. D. F., and Joyce, C. M. (1992) *J. Biol. Chem.* 267, 8417–8428
11. Steitz, T. A. (1993) *Curr. Opin. Struct. Biol.* 3, 31–38
12. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature* 391, 251–258
13. Donlin, M. J., Patel, S. S., and Johnson, K. A. (1991) *Biochemistry* 30, 538–546
14. Johnson, S. J., Taylor, J. S., and Beese, L. S. (2003) *Proc. Natl. Acad. Sci. U. S. A.* 100, [illegible]
15. Gibbs, P. R., and [illegible] (19[illegible]) *[illegible]*

497–504.
36. Kong, H., Raghunathan, P. C., and Wang, T. S.-F. (1993) *J. Biol. Chem.* **268**, 1965–1975.
37. Kong, H. M., Sousad, S. G., Copper, V. L., and Goodman, S. d. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 2278–2282.
38. Hughes, S. H., Ellenberger, T., Hogle, N. A., Leavy, P., Arnheiter, M., Stover, S., and Sagerman, P. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 5992–5996.
39. Shaw, V., Jacobsen, G., Maxwell, T., Laughter, L., Lawless, R., and Rodgers, J. (1990) *Biotechniques* **7**, 1180–1182.
40. Sweasy, A., Lawson, A. M., Shank, L., and Sakai, H. (1990) *J. Mol. Biol.* **214**, 753–761.
41. Strategene, V., Bloom, L., and Sakai, H. (1990) *J. Mol. Biol.* **214**, 67–81.
42. Rodriguez, A. C., Park, H.-W., Mao, C., and Beese, L. S. (2007) *J. Mol. Biol.* **299**, 447–462.
43. Franklin, M. C., Wang, J., and Steitz, T. A. (2001) *Cell* **105**, 657–667.
44. Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., and Kai, Y. (2001) *J. Mol. Biol.* **306**, 469–477.
45. Yang, G., Franklin, M., Li, J., Lin, T.-C., and Konigsberg, W. (2002) *Biochemistry* **41**, 8546–8556.
46. Astatke, M., Ng, K., Grindley, N. D. F., and Joyce, C. M. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3402–3407.
47. Astatke, M., Grindley, N. D. F., and Joyce, C. M. (1998) *J. Mol. Biol.* **278**, 147–165.
48. Brandis, J. W. (1988) *Nucleic Acids Res.* **16**, 1013–1035.
49. Kong, H. S., and Kucera, R. B. (1990) *Gene (Amst.)* **92**, A1–A4 (Addenda).
50. Yang, G., Franklin, M., Li, J., Lin, T.-C., and Konigsberg, W. (2002) *Biochemistry* **41**, 10550–10557.
51. King, K., Benner, S. E., and Cook, W. E. (1992) *J. Biol. Chem.* **268**, 21661–21670.
52. Gardner, A. F., and Jack, W. E. (1999) *Nucleic Acids Res.* **27**, 2545–2553.
53. Spanner, A. F., and Jack, W. E. (2002) *Nucleic Acids Res.* **30**, 5010–5018.
54. Rever, S. J., Fogg, M. J., Skorvaga, M., Beese, N., Pence, D. H., and Connolly, B. A. (2002) *Nucleic Acids Res.* **30**, 1009–1018.
55. Parks, C. B., Hunter, S., and King, K. (1998) *Gen. Physiol. Chem.* **40**, 277–286.
56. Jones, D., Stevens, C. J., and Hagerman, M. H. (2003) *J. Mol. Biol.* **325**, 709–720.
57. Jones, L. N., Wynne, C. N., Schweiger, G., Cavaluzzi, T., Estes-Gilreath, H., Bushey, M., Sutherland, K., and Blocker, H. F. (1999) *Biotechniques* **26**, 866–872.
58. Pearson, M. G. (1999) *Nucleic Acids Res.* **19**, 3013.
59. Sandler, A. F. (1999) *Methods Mol. Biol.* **96**, 63–80.
60. Fisher, J. M., Fraser, G. L., Pace, N. J., Hughes, C. W., Robertson, C. N., Rajagopalan, B. J., Carreira, A. O., Jackson, M. A., and Stevenson, R. (1987) *Science* **235**, 329–331.
61. Rhee, S., Jarvis, T., and Beese, L. S. (2000) *Genome Res.* **9**, 922–932.
62. Beese, L. S., Derbyshire, V., Steitz, T. A. (2001) *Science* **291**, 774–779.
63. Hogrefe, H. H., Hansen, C. J., and Cline, J. (2001) *Nucleic Acids Res.* **29**, 2354–2360.
64. Thierry, G. L. (February 7, 1995) U.S. Patent 5,387,661.
65. Loeb-Kolin, C., Townsend, M. S., Sanderson, M. R., Joyce, L., Kirchner, J. M., Greco, C. M., and Smith, T. S. (2002) *Biotech. Bull.* **8**, 180–203.
66. Johnson, K. A. (1993) *Methods Enzymol.* **249**, 38–61.
67. Perrin, A. S., Nti, E. P., Kuchibhotla, J., Lowe, B. M., Williams, A. J., Sun, D. M., Black, S., Carter, P., Wyss, M. N., and Wilmer, R. (2000) *Anticancer Res.* **20**, 2253–2261.
68. Goodman, S. J., and Concorde, C. S. (2000) *Methods Enzymol.* **262**, 557–568.
69. Tabor, S., and Richardson, C. C. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8305–8342.
70. Nicolas, D. T., Lee, L. Greeley, N. D. P., Rotnick, T. A., and Alpar, C. M. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1159–1168.

doi:10.1006/jmbi.2000.4488 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (2001) 308, 485–496

JMB

Crystal Structure of DNA Polymerase from Hyperthermophilic Archaeon *Pyrococcus kodakaraensis* KOD1

Hiroshi Hashimoto[1], Motomu Nishioka[2], Shinsuke Fujiwara[2] Masahiro Takagi[2], Tadayuki Imanaka[2], Tsuyoshi Inoue[3] and Yasushi Kai[3]*

[1]*Department of Materials Chemistry and*

[2]*Department of Biotechnology Graduate School of Engineering, Osaka University 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan*

[3]*Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University Yoshidahonmachi Sakyo-ku Kyoto 606-8501, Japan*

*Corresponding author

The crystal structure of family B DNA polymerase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 (KOD DNA polymerase) was determined. KOD DNA polymerase exhibits the highest known extension rate, processivity and fidelity. We carried out the structural analysis of KOD DNA polymerase in order to clarify the mechanisms of these enzymatic features. Structural comparison of DNA polymerases from hyperthermophilic archaea highlighted the conformational differences in Thumb domains. The Thumb domain of KOD DNA polymerase shows an "opened" conformation. The fingers subdomain possessed many basic residues at the side of the polymerase active site. The residues are considered to be accessible to the incoming dNTP by electrostatic interaction. A β-hairpin motif (residues 343–349) extends from the 3′-exonuclease (Exo) domain as seen in the editing complex of the RB69 DNA polymerase from bacteriophage RB69. Many arginine residues are located at the forked-point (the junction of the template-binding and editing clefts) of KOD DNA polymerase, suggesting that the basic environment is suitable for partitioning of the primer and template DNA duplex and for stabilizing the partially melted DNA structure in the high-temperature environments. The stabilization of the melted DNA structure at the forked-point may be correlated with the high PCR performance of KOD DNA polymerase, which is due to low error rate, high elongation rate and processivity.

© 2001 Academic Press

Keywords: archaea; crystal structure; family B DNA polymerase; "forked-point"; KOD DNA polymerase

Introduction

DNA polymerases are a group of enzymes that use single-stranded DNA as a template for the synthesis of the complementary DNA strand. These enzymes are multifunctional, with both synthetic (polymerase) and one or two degradative modes (5′-3′ and/or 3′-5′ exonucleases) and play an essential role in nucleic acid metabolism including the processes of DNA replication, repair and recombination. Many DNA polymerase genes have been cloned and sequenced. Amino acid sequences deduced from their nucleotide sequences can be classified into four major types: Escherichia coli DNA polymerase I (family A), E. coli DNA polymerase II (family B), E. coli DNA polymerase III (family C) and others (family X).[1] Recently, a new family of DNA polymerases has been identified. All members of this family contain five highly conserved motifs, I–V, and several of these polymerases participate in lesion bypass.[2] This family is called the Dinb/UmuC/DinB family.[3] Family B DNA polymerases include eukaryotic DNA polymerases α, δ, and ε, which are thought to be components of the replisome and to carry out chromosomal DNA replication. Archaeal proteins involved in gene expression, such as those for DNA replication, transcription, and translation, have been found to be similar to those from eucarya. Therefore, the archaeal system of gene expression is a simplified model of the eukaryotic system. In contrast, the

E-mail address of the corresponding author: kai@chem.eng.osaka-u.ac.jp

0022-2836/01/020485–9 $35.00/0

cellular appearance and organization of archaea are more like those of bacteria.

The first crystal structure of a family B DNA polymerase to be obtained was that of bacteriophage RB69 DNA polymerase (RB69 DNA polymerase).[4] The first crystal structure of archaeal DNA polymerase was DNA polymerase from *Thermococcus gorgonarius* (Tgo DNA polymerase).[5] The editing complex of RB69 DNA polymerase has been reported.[6] Two further crystal structures of archaeal family B DNA polymerases have recently been reported: Tok DNA polymerase from *Desulfurococcus* sp. Tok[7] at 9°N-7 DNA polymerase from *Thermococcus* sp. 9°N-7.[8]

The *Pyrococcus kodakaraensis* KOD1 is a hyperthermophilic archaeon, with an optimum growth temperature of 95°C.[9] Enzymes produced in KOD1 were reported to be extremely thermostable and to have eukaryotic characteristics.[9] The optimum temperature of KOD DNA polymerase is 75°C, similar to that of DNA polymerase obtained from *Pyrococcus furiosus* (Pfu DNA polymerase). KOD DNA polymerase, however, exhibits the highest extension rate (100–130 nucleotides/second) and processivity (>300 bases); five times and ten to 15 times higher than those of Pfu DNA polymerase, respectively.[10] Thermostable DNA polymerases are expected to be suitable enzymes for Polymerase Chain Reaction (PCR). KOD DNA polymerase is, therefore, suitable for DNA amplification by such means. Indeed, KOD DNA polymerase is widely used in rapid and accurate PCR systems (TOYOBO Ltd., Japan).

Although structures of three archaeal DNA polymerases have been determined as described above, no structural information relating to elongation rate, processivity or fidelity is provided. We carried out the structural analysis of KOD DNA polymerase in order to clarify the mechanism of enzymatic features of KOD DNA polymerase, which are the highest extension rate, processivity and fidelity. Here, we report the crystal structure of DNA polymerase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. The three-dimensional structure of this KOD DNA polymerase may provide useful information to clarify the mechanisms for rapid and accurate reaction. In addition, this information may contribute to the improvement of the PCR properties of enzymes already in use such as thermostability, error rate, elongation rate and processivity, or for designing new enzymes for PCR as well as DNA replication by family B DNA polymerases.

Results and Discussion

Overall structure

KOD DNA polymerase has a disk-like shape with dimensions 68 Å × 80 Å × 100 Å and is made up of distinct domains and subdomains: N-terminal (N-ter; 1–138, 327–358, violet), Exonuclease (Exo; 139–326, blue), Polymerase (Pol) domain including the Palm and Fingers subdomains (359–449, 580–687, brown) and 450–499, green, respectively) and the Thumb domain including thumb-1 and thumb-2 subdomains (688–774, red) (Figure 1(a)). The polymerase active site, consisting three conserved carboxylates (Asp404, Asp540 and Asp542) is located in an anti-parallel β-sheet in the Palm subdomain. The exonuclease active site contains two conserved carboxylates (Asp141 and Glu143) and is located in an anti-parallel β-sheet in the Exo domain. The Polymerase and exonuclease active sites on the molecular surface are indicated by P and E, respectively (see Figure 4). Structural comparisons of archaeal DNA polymerases (KOD, Tgo and 9°N-7 DNA polymerases) are shown in Figure 1(b). The structural architectures of the proteins are identical, but the orientation of the domains and subdomains is different. In the case of the KOD DNA polymerase (red), the Thumb domain is shifted to make an "open" conformation and the portion of the Palm domain neighboring the root of the Thumb domain is slightly shifted as a result of the large movement of the Thumb domain in comparison to other archaeal DNA polymerases. Table 1 shows the averaged temperature factors of the domains and subdomains in the crystal structure of KOD DNA polymerase. The value of the Thumb domain was markedly higher than the others. The structures of many residues in the Thumb-2 subdomain are not defined, because the orientation of the subdomain is highly disordered. Therefore, it is thought that the structure of KOD DNA polymerase described here provides information for the DNA-free, most relaxed conformation. The structure of the editing complex of RB69 DNA polymerase revealed that newly synthesized duplex DNA is grasped by the Pol and Thumb domains. Although the orientation of the Thumb domain is potentially highly flexible, the orientation may be fixed when it binds to the primer-template duplex.

Polymerase domain

The Pol domain is made up of the Fingers and Palm subdomains and has an "L-like" shape (Figure 2(a)). The polymerization mechanism has been studied mainly on family A DNA polymerases (Pol I). A structural basis for a motch-

Table 1. Averaged temperature factors

Domain	Temperature factor (Å²)
N-ter	38.1
Exo	39.7
Pol	
Fingers	49.5
Palm	52.8
Thumb	58.7
Overall	50.9

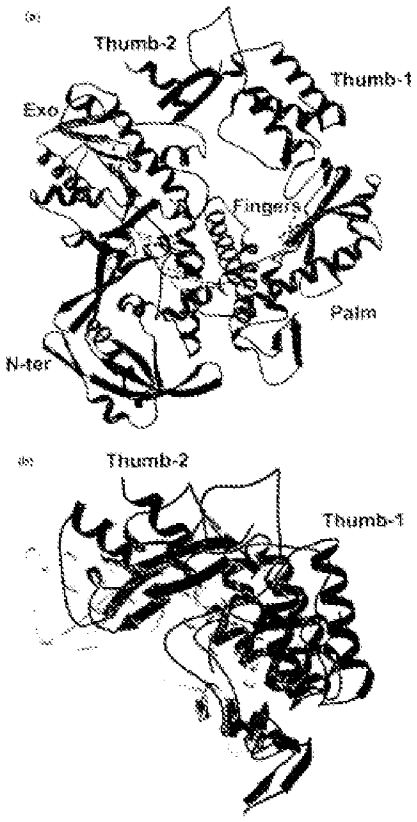


Figure 3. (a) Overall structure of KOD DNA polymerase. The structure is composed of domains and subdomains, which are N-terminal (N-ter, violet), Exonuclease (Exo, blue), Polymerase (Pol) domain including the Palm (brown) and Fingers (green) subdomains and the Thumb domain (red), including the Thumb-1 and Thumb-2 subdomains. Conserved carboxylate residues in Polymerase and Exonuclease active site are shown by ball-and-stick models. (b) Conformational comparison of Thumb domains among three archaeal DNA polymerases. Red, KOD DNA polymerase; blue, Tgo DNA polymerase and green, 9°N-7 DNA polymerase. The comparison shows that the Thumb domain of KOD DNA polymerase displays the most "opened" conformation.

assisted mechanism of phosphoryl transfer was provided by the bacteriophage T7 DNA replication complex.[1] The complex structure shows that two metal ions are bound by strictly conserved carboxylates (Asp475 and Asp656, which correspond to Asp480 and Asp542 in KOD DNA polymerase) extended from the anti-parallel β-sheet of the Palm domain. The phosphate group of incoming dsCTP is held by the metal ions and the four basic residues extending from the Fingers subdomain (His506, Arg518 and Lys520). The crystal structure of two ternary complexes of the large fragment of

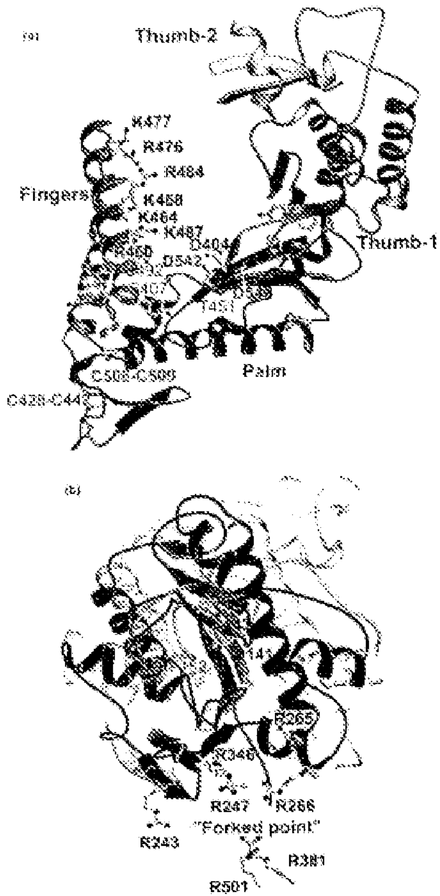


Figure 2. (a) Ribbon representation of the Pol domain. The domain is made up of Fingers and Palm subdomains. Conserved carboxylate residues (D404, D540 and D542) are represented by ball-and-stick models. Basic residues are represented by ball-and-stick models, which stand in a line in the Fingers subdomain, facing the polymerase active site. K484 are replaced by alanine, because of the ambiguity of its electron density. Two disulfide bonds are displayed (C428-C442 and C506-C509). Aromatic residues adjacent to a glycine residue, represented by ball-and-stick models, are localized in the [illegible] of the subdomains. The Thumb domain is represented by the semitransparent model. The Cα atoms of the nucleophilic residues, S407 (Pfu Pol-I), S492 (Tli Pol-I) and [illegible] (Tli Pol-2), are represented by violet spheres. (b) Exonuclease domains of KOD DNA polymerase and RB69 DNA polymerase (semitransparent model). Conserved carboxylate (D141 and E143) and arginine residues (R243, R247, R265, R266, R345, R381 and R501) at the forked-point of KOD DNA polymerase are represented by ball-and-stick models. The red strands are β-hairpin motif partitioning template-binding and editing clefts. The loop containing Phe152 is shown in orange. F123 of RB69 DNA polymerase and F152 of KOD DNA polymerase are represented by semitransparent and opaque ball-and-stick models, respectively.

Thermus aquaticus DNA polymerase I (Klentaq); with a primer-template DNA and ddCTP have been reported.[12] The ternary complexes suggest that basic residues at the fingers subdomain hold the phosphate group of the incoming dNTP and the domain induces a conformational change to deliver the incoming nucleotide to the active site. In the case of family B DNA polymerases, the fingers subdomain is composed mainly of rare long helices and does not have a joint that appeared in the structures of family A DNA polymerases. Therefore, it seems that in the case of archaeal DNA polymerases, the movement of the Exo domain to deliver dNTP to the active site differs from that of family A DNA polymerases. Kinetic study of RB69 DNA polymerase mutants revealed that four residues (Arg482, Lys486, Lys560 and Asn561) of the fingers subdomain affected dNTP incorporation.[13] The residues are conserved in family B DNA polymerases, and correspond to Arg460, Lys464, Lys487 and Asn488 in KOD DNA polymerase, respectively. Furthermore, Lys468, Arg476, Lys477 and Arg484 are located at the tip of the fingers subdomain on the side of the polymerase active site in KOD DNA polymerase (Figure 2(a)). It is expected that the "queue" of basic residues captures the incoming dNTPs, then the dNTP is delivered toward the polymerase active-site under by accompanying the movement of the polymerase domain. Two disulfide bonds exist in the connection site between the Palm and Fingers subdomains (Figure 2(a): Cys428-Cys482 and Cys506-Cys509). The two disulfide bonds are found also in the crystal structures of Tgo, Tok and 9°N-7 DNA polymerases. Sequence alignment for archaeal DNA polymerases is shown in Figure 3, suggesting the potential for the formation of disulfide bonds in the same sites. It is thought that the disulfide bonds are required to maintain the structure of the Fingers and Palm subdomains at extremely high temperatures. Sequence comparison suggests that the number of disulfide bonds are correlated with optimum growth temperatures of organisms. DNA polymerases from *Thermococcus litoralis*, *Methanococcus jannaschii* and *Archaeoglobus fulgidus*, with optimum growth temperatures of 88, 85 and 83 °C, respectively, are expected to have one disulfide bond, because Cys506 is replaced by serine in *T. litoralis* and *M. jannaschii*, and Cys482 is replaced by arginine in *A. fulgidus*. DNA polymerase from *Methanobacterium thermoautotrophicum*, with an optimum growth temperature of 65 °C, is expected to have no disulfide bond, because Cys428, Cys482 and Cys506 are replaced by glutamic acid, arginine and serine, respectively.

Archaeal DNA polymerases have characteristic sequences of aromatic residues adjacent to glycine residues (Figure 3). These are located at the hinges of the Palm subdomain at the connections to the Fingers and Thumb-1 subdomains (Figure 2(a)). These aromatic residues may provide a flexible aromatic environment because of the adjoining glycine residues. This may contribute to the conformational changes of Pol domain in polymerization.

The 3′-5′ exonuclease domain

DNA is synthesized by competition between the rate of polymerase and exonuclease activities at the newly synthesized 3′ terminus from the primer. Misincorporation of a nucleotide destabilizes the structure of duplex DNA at the 3′ terminus of the primer. This decreases the rate of nucleophilic attack on the orthophosphate group of the incoming dNTP by the primer 3′-OH and allows excision of the incorrect nucleotide by the proofreading exonuclease. The excision requires the movement of the 3′ terminus to the exonuclease active site accompanied by rewinding of the duplex DNA, because the exonuclease active site is set apart from the polymerase active site. In KOD DNA polymerase, the exonuclease active site is set apart from the polymerase active site by approximately 40 Å. The editing complex of RB69 DNA polymerase shows structural similarity to the editing mode of family B DNA polymerase.[4] The DNA polymerase binds the mismatched primer-template DNA, where is partially denatured, the 3′ end of the primer strand is bound at the exonuclease site. Residues 253–262 of RB69 DNA polymerase, first form an extended β-hairpin structure that juts directly out from the protein surface and projects into the DNA, stabilize the partially denatured or melted structure. Arg258 extending from the β-hairpin motif plays an important role. Arg258 and Phe123 appear to block the template strand by making interactions with the penultimate base at the 3′ end of the primer-template. Arg258 and Phe123 in RB69 DNA polymerase correspond to Arg257 and Phe152, in KOD DNA polymerase respectively. Figure 2(b) shows the structural comparison of Exo domains of KOD and RB69 DNA polymerases. Molecular surface and electrostatic potentials are shown in Figure 4. The β-hairpin motif in KOD DNA polymerase corresponds to residues 247–249 and Arg247, extending to the forked-point, which is the junction of the template-binding and editing clefts (T-cleft and E-cleft, respectively) (Figure 4). It seems that Arg247 can separate template strand from primer strand and stabilize the melted structure of the strands in a manner similar to that of the RB69 DNA polymerase. As Phe152 is set apart from the active site, it is apparently unable to make an aromatic interaction with the base of the primer. Based on the above idea, the movement of the loop including Phe152 (Figure 2(b)) is required to interact with the primer strand at the E-cleft. Furthermore, Arg243 extends from the β-hairpin structure to the T-cleft. Arg243 interacts with the template strand to fix it at the T-cleft. In addition to Arg243 and Arg247, five arginine residues gather at the forked-point in KOD DNA polymerase (Arg245, Arg246, Arg248, Arg382 and Arg381) and provide a basic environment (Figures 2(b) and 4). It seems that they can interact with the phosphate

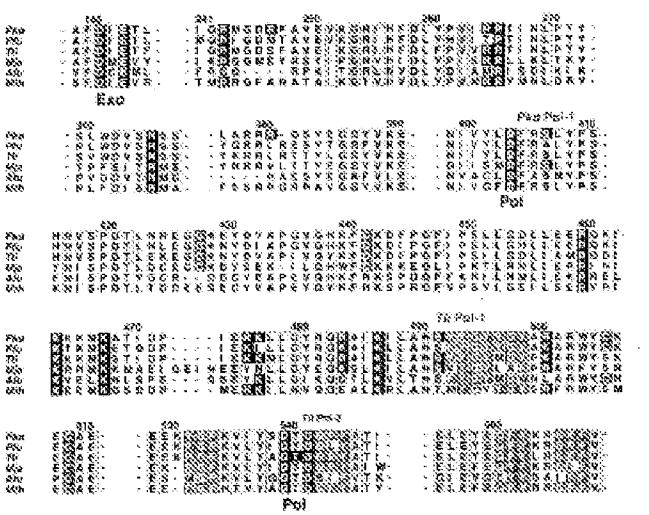


Figure 3. Sequence alignment of archaeal DNA polymerases. The abbreviations used are as follows: Pfu, *Pyrococcus furiosus*; Pho, *Pyrococcus horikoshii*; Tli, *Thermococcus litoralis*; Mja, *Methanococcus jannaschii*; Afu, *Archaeoglobus fulgidus*; and Mth, *Methanobacterium thermoautotrophicum*. Homologous residues are numbered in gray. Exonuclease residues are highlighted in reverse type. Conserved carboxylate residues in the Exonuclease and Polymerase active sites are shown in red. Basic residues gathering in the forked-point and Finger subdomain are shown in blue. R345, R347, S355, R368, R386, R391 and R393 are located in the forked-point. R460, R464, R468, R476, R477, R485 and R487 are located in the Finger subdomain and face into the polymerase active site. Cysteine residues forming (or possibly forming) disulfide bonds are shown in green. Nucleophilic residues in self-splicing reaction are shown in violet. Introns interrupt before the nucleophilic residues. Aromatic residues adjacent to glycines are shown in orange.

groups of the DNA strand and stabilize the melted structure of DNA strands at the forked-point. Several arginine residues at the forked-point are conserved in known family B DNA polymerases from hyperthermophilic archaea.

In DNA synthesis, the structure of DNA is variable at the stage of switching between the elongation and editing modes. Hyperthermophiles must have mechanisms to protect their genomic DNA against thermal denaturation. The genomic DNA of hyperthermophilic archaea have nucleosome-like structures brought about by interaction with histone-like proteins.[13] Nevertheless, at the replication fork, the DNA strands are exposed. Therefore, DNA polymerases of hyperthermophilic archaea are required to stabilize the exposed or melted DNA structure in the high temperature environment. The stabilization by DNA polymerase may correlate with the enzymatic characteristics of DNA polymerase such as half-life period of activity, error rate, elongation rate and processivity. As discussed above, it is considered that the arginine residues around the "forked-point" have a remarkable effect on the stability of DNA structure. In the forked-point of *Pfu* DNA polymerase, Arg247, Arg285 and Arg301 are replaced by methionine, threonine and lysine, respectively. Therefore, the replacements may affect the difference of the enzymatic characteristics between KOD and *Pfu* DNA polymerases. Additional experiments such as site-directed mutagenesis, together with enzymatic studies of DNA polymerases are necessary to clarify the role of the residues at the forked-point.

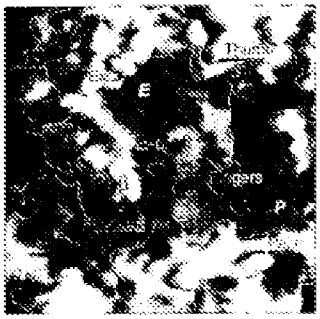


Figure 4. Molecular surface with electrostatic potential map around the bound primer. The red and blue surfaces are acidic and basic regions, respectively. Exonuclease and subdomains are labeled with orange letters. Polymerase and Exonuclease active sites are labeled with P and E, respectively. The β-hairpin is labeled with β.

Extein connection site

The KOD DNA polymerase gene encodes a 3673 amino acid residue precursor protein. The precursor protein is processed precisely into three parts by protein splicing. The self-splicing reaction yields the mature KOD polymerase (774 residues) and two intervening protein domains (termed inteins), Pfu-Pol (360 residues) and Pfu-Pol2 (537 residues) as a result of the ligation of the external N and C-terminal domains (termed exteins).[10,15] All known precursor proteins contain conserved serine amino acid self-splicing sites: serine, threonine or cysteine (nucleophiles) at the intein N terminus, and His-Asn pair at the intein C terminus followed by serine, threonine, cysteine (nucleophiles) at the C-extein N-terminus.[15] The traces of the protein splicing reaction in KOD DNA polymerase are Ser407 and Ser932, which were located in at the N terminus of the C-extein in the crystal structure of KOD DNA polymerase; the nucleophilic residues are found in the Pol domain (Figure 3(a)).

Self-splicing sites in archaeal family B DNA polymerases (a family) are classified into three types: *Pfu* Pol-1, *Tli* Pol-3 and *Tli* Pol-2 (The Intein Database, http://www.neb.com/neb/inteins.html). The nucleophilic residues, serine or threonine, in the three sites are mapped in Figure 3(a). In the case of KOD DNA polymerase, Pfu-Pol2 intervenes in the *Pfu* Pol-1 site and Pfu-Pol1 intervenes in the *Tli* Pol-1 site. The structure shows that they are localized around the polymerase active site in the Palm domain. Although they are exposed to solvent, they are surrounded by the Fingers subdomain and the Thumb domain. The two inteins cannot exist in the space because of steric hindrance. Therefore, it is necessary that the folding of inteins and the subsequent self-excisions are carried out before the extein is folded.

Materials and Methods

Crystallization

KOD DNA polymerase was overexpressed in *E. coli* BL21(DE3) and purified by the previously reported method.[10] The crystals of KOD DNA polymerase were grown by the previously reported method.[10] KOD DNA polymerase was concentrated up to about an $A_{280\ nm}$ of 25. Crystals of KOD DNA polymerase suitable for diffraction experiments were obtained at 288 K with hanging drops of 2 μl of protein solution and 2 μl of reservoir solution, containing 100 mM sodium citrate buffer (pH 5.6) and 25–30% (v/v) 2-methyl-2,4-pentanediol (MPD), equilibrated against the reservoir solution.

Data collection

X-ray diffraction measurements were performed at the beamline 18B of the Photon Factory at the High Energy Accelerator Research Organization, Tsukuba Science City, Japan. Each crystal of KOD DNA polymerase was picked up directly with a nylon fiber loop from a drop of mother liquid, the crystal was then rapidly transferred to the N_2 gas stream. The incident beam with wavelength of 1.08 Å was collimated to 0.2 mm in diameter. Intensity data were collected on 300 mm × 300 mm imaging plates (Fuji Film Company Ltd.) using the Weissenberg camera for macromolecules with a radius of 450 mm[16,17] and the oscillation method with 3° rotation per frame. The crystals diffracted to at least 2.8 Å resolution at 100 K. X-ray diffraction data were processed and scaled with programs DENZO and SCALEPACK.[18] The diffraction data were scaled with zero σ cutoff. Unit-cell parameters were determined as $a = 131.9$ Å, $b = 132.6$ Å, and $c = 79.0$ Å with the space group of $P2_12_12_1$. The unit-cell parameters gave Matthews's coefficient of 3.60 Å^3 Da^{-1} and a solvent content of 52.2% (v/v).[19] The final completeness of the data consisted of 124,888 measurements of 38,788 unique observed reflections with an overall R_{merge} of 8.6% and 34.8% in the outermost resolution shell (2.86–2.80 Å). This represents 98.1% of theoretically observable reflections at 2.8 Å resolution. The outermost resolution shell of data is 85.7% complete.

Structure determination

The crystal structure of KOD DNA polymerase was solved by molecular replacement with the AMoRe program.[20] The structure of Tgo DNA polymerase (PDB code 1TGO) reduced to polyalanine was used as the search model. Data in the resolution range of 30.0–3.5 Å were used to find the rotation and translation functions. Results are discussed in terms of the AMoRe correlation coefficient (CC). Using a Patterson cut-off radius of 35 Å, a list of 20 rotation function peaks was obtained, with the top peak having an AMoRe CC value of 13.8. The top solution by translation function is CC of 43.3 with an R-factor of 54.1%. At this stage, the electron density of the Thumb domain is very ambiguous. Therefore, structural refinement of the initial stage was carried out with

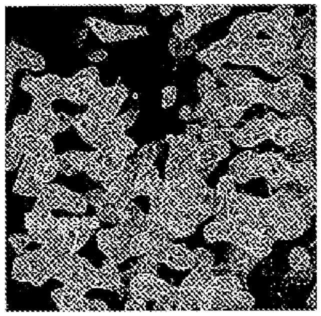


Figure 5. The final $2F_o - F_c$ map around the Fingers and Palm subdomains. The map is contoured at 1 σ.

a model omitting the Thumb domain. The model was manually modified using the program O[22] and subjected to further rounds of refinement using data in the resolution range 40.0-3.0 Å with the program CNS.[24] The final R-factor is 22.1% and R_{free} is 31.3%, with r.m.s. deviations for bond lengths and bond angles being 0.007 Å and 1.1°, respectively. The 58 residues at the tip of one Thumb domain are not included in the final model due to poorly defined electron density. Figure 5 shows the final $2F_o - F_c$ map superimposed on the refined final coordinates of KOD DNA polymerase.

Protein Data Bank accession code

Refined coordinates and structure factor have been deposited in the RCSB Protein Data Bank under the accession code 1GCX.

Figure preparation

Figures 1 and 2 were prepared using programs MOLSCRIPT[27] and Raster3D.[28,29] Figure 4 was prepared by GRASP.[30] Figure 5 was prepared using the program O.[22]

Acknowledgments

We thank Professor N. Sakabe, Dr N. Watanabe, Dr M. Suzuki and Dr N. Igarashi, for support in data collection at KEK-PF, Japan. This study was supported by TARA Sakabe Project of University of Tsukuba. The author is grateful for a JSPS Fellowship for Japanese Junior Scientists.

References

1. Braithwaite, D. K. & Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucl. Acids Res. 21, 787-802.
2. Johnson, R. E., Washington, M. T., Prakash, S. & Prakash, L. (1999). Bridging the gap: a family of novel DNA polymerases that replicate faulty DNA. Proc. Natl Acad. Sci. USA, 96, 12224-12226.
3. Friedberg, E. C., Feaver, W. J. & Gerlach, V. L. (2000). The many faces of DNA polymerases: strategies for mutagenesis and for mutational avoidance. Proc. Natl Acad. Sci. USA, 97, 5681-5683.
4. Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997). Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. Cell, 89, 1087-1099.
5. Hopfner, K.-P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R. & Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius. Proc. Natl Acad. Sci. USA, 96, 3600-3605.
6. Shamoo, Y. & Steitz, T. A. (1999). Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. Cell, 99, 155-166.
7. Zhao, Y., Jeruzalmi, D., Moarefi, I., Leighton, L., Lasken, R. & Kuriyan, J. (1999). Crystal structure of an archaebacterial DNA polymerase. Structure, 7, 1189-1199.
8. Rodriguez, A. C., Park, H.-W., Mao, C. & Beese, L. S. (2000). Crystal structure of a pol α family DNA polymerase from the hyperthermophilic archaeon Thermococcus sp. 9°N-7. J. Mol. Biol. 299, 471-487.
9. Fujiwara, S., Okuyama, S. & Imanaka, T. (1996). The world of archaea: genome analysis, evolution and thermostable enzymes. Gene, 179, 165-170.
10. Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M. & Imanaka, T. (1997). Characterization of DNA polymerase from Pyrococcus sp. strain KOD1 and its application to PCR. Appl. Environ. Microbiol. 63, 4504-4510.
11. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. Nature, 391, 251-258.
12. Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. EMBO J. 17, 7514-7525.
13. Yang, G., Lin, T.-C., Karam, J. & Konigsberg, W. H. (1999). Steady-state kinetic characterization of RB69 DNA polymerase mutants that affect dNTP incorporation. Biochemistry, 38, 8094-8101.
14. Grayling, R. A., Sandman, K. & Reeve, J. N. (1996). DNA stability and DNA binding proteins. Advan. Protein. Chem. 48, 437-467.
15. Nishioka, M., Fujiwara, S., Takagi, M. & Imanaka, T. (1998). Characterization of two intein homing endonucleases encoded in the DNA polymerase gene of Pyrococcus kodakaraensis strain KOD1. Nucl. Acids Res. 26, 4409-4412.
16. Perler, F. B., Olsen, G. J. & Adam, E. (1997). Compilation and analysis of intein sequences. Nucl. Acids Res. 25, 1087-1093.
17. Hashimoto, H., Matsumoto, T., Nishioka, M., Yuasa, T., Takeuchi, S., Inoue, T., Fujiwara, S., Takagi, M.,

USSN 10/089,027
Atty. Docket No. 2007651-0001

4567604v1

Imanaka, T. & Kai, Y. (1999). Crystallographic studies on a family B DNA polymerase from hyperthermophilic archaeon Pyrococcus kodakaraensis strain KOD1. J. Biochem. 125, 983-986.

18. Sakabe, N., Ikemizu, S., Sakabe, K., Higashi, T., Nakagawa, A., Watanabe, N., Adachi, S. & Sasaki, K. (1995). Weissenberg camera for macromolecules with imaging plate data collection system at the photon factory; present status and future plan. Rev. Sci. Instrum. 66, 1276-1281.
19. Watanabe, N., Nakagawa, A., Adachi, S. & Sakabe, N. (1995). Macromolecular crystallography station BL-18B at the photon factory. Rev. Sci. Instrum. 66, 1824-1826.
20. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307-326.
21. Matthews, B. W. (1968). Solvent content of protein crystals. J. Mol. Biol. 33, 491-497.
22. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Crystallog. sect. A, 50, 157-163.
23. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in the model. Acta Crystallog. sect. A, 47, 110-119.
24. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Crystallog. sect. D, 54, 905-921.
25. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallog. 24, 946-950.
26. Merritt, E. A. & Murphy, M. E. (1994). Raster3D version 2.0: a program for photorealistic molecular graphics. Acta Crystallog. sect. D, 50, 869-873.
27. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. Methods Enzymol. 277, 505-524.
28. Nicholls, A., Sharp, K. A. & Honig, B. (1991). GRASP: Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins: Struct. Funct. Genet. 11, 281-296.

Edited by R. Huber

(Received 22 August 2000; received in revised form 8 December 2000; accepted 8 December 2000)

Crystal structure of an archaebacterial DNA polymerase

Yanxiang Zhao[1†], David Jeruzalmi[1†], Ismail Moarefi[1,2], Lore Leighton[1,2], Roger Lasken[3] and John Kuriyan[1,2*]

Background: Members of the Pol II family of DNA polymerases are responsible for chromosomal replication in eukaryotes, and carry out highly processive DNA replication when attached to ring-shaped processivity clamps. The sequences of Pol II polymerases are distinct from those of members of the well-studied Pol I family of DNA polymerases. The DNA polymerase from the archaebacterium *Desulfurococcus strain* Tok (D. Tok Pol) is a member of the Pol II family that exhibits catalytic activity at elevated temperatures.

Results: The crystal structure of D. Tok Pol has been determined at 2.8 Å resolution. The architecture of this Pol II type DNA polymerase resembles that of the DNA polymerase from the bacteriophage RB69, with which it shares less than ~20% sequence identity. As in RB69, the central catalytic region of the DNA polymerase is located within the 'palm' subdomain and is strikingly similar in structure to the corresponding regions of Pol I type DNA polymerases. The structural scaffold that subtends the catalytic core in D. Tok Pol is consistent in structure to that of Pol I type polymerases. The 3′–5′ proofreading exonuclease domain of D. Tok Pol resembles the corresponding domains of RB69 Pol and Pol I type DNA polymerases. The exonuclease domain in D. Tok Pol is located at the same position relative to the polymerase domain as seen in RB69, and on the opposite side of the palm subdomain compared to its location in Pol I type polymerases. The 5′ terminal domain of D. Tok Pol has structural similarity to DNA-binding proteins. Sequence alignments suggest that this domain is conserved in the eukaryotic DNA polymerases δ and ε.

Conclusions: The structure of D. Tok Pol confirms that the modes of binding of the template and extension of newly synthesized duplex DNA are likely to be similar in both Pol II and Pol I type DNA polymerases. However, the mechanism by which the newly synthesized product strand is sent out of the proofreading exonuclease domain has to be quite different. The discovery of a domain that seems to be an RNA-binding module raises the possibility that Pol II family members interact with RNA.

Addresses: [1]Laboratories of Molecular Biophysics, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA and [2]Molecular Biology, Inc., 66 High Street, Oxford, OH 45056, USA.

[†]The two authors contributed equally to this work.

*Corresponding author.
E-mail: kuriyan@rockvax.rockefeller.edu

Key words: Desulfurococcus Tok, DNA polymerase, Pol II family, RNA-binding domain, thermostability, X-ray crystallography

Received: 12 April 1999
Revised received: 27 May 1999
Published: 8 June 1999
Accepted: 8 June 1999

Published: 29 September 1999

Structure October 1999, 7:1189–1199
http://biomednet.com/elecref/0969212600701189

Introduction

DNA polymerases can be classified into at least three families on the basis of sequence similarities in the three dimensional DNA polymerases of *Escherichia coli*, Pol I, Pol II and Pol III [1–3]. Members of the Pol I family have been studied extensively, leading to a comprehensive understanding of their functional properties and their structure [3–6]. In contrast to the detailed knowledge that is now available for the Pol I family, the Pol II and Pol III polymerases are poorly understood. The first crystal structure determined for a Pol II family member was that of the DNA polymerase of the bacteriophage RB69 (RB69 Pol) [7] and no structural information is currently available for any member of the Pol III family. Members of the Pol II family known as Pol B or Pol α) and Pol III families carry out processive replication of chromosomal DNA during cell division [8], and there is interest in further extending our knowledge of their structures and mechanisms. Archaebacterial DNA polymerases and the eukaryotic DNA polymerases α, δ and ε are members of the Pol II family [1].

The structure of RB69 Pol revealed that the general architecture of the core of the Pol II polymerases is strikingly similar to that of the Pol I polymerases [7]. Pol I polymerases are economized from three smaller subdomains, termed the thumb, palm and fingers regions by analogy to elements first noted in the structure of the Klenow fragment of *E. coli* DNA polymerase I [9]. In addition, Pol I DNA polymerases have a proofreading 3′–5′ exonuclease domain located below the thumb subdomain, near the region where duplex DNA exits the polymerase active site [8,5]. Besides the residues involved in catalysis, there is no significant sequence similarity between the polymerase domains of members of the Pol I and Pol II families [1].

However, the subdomain architecture of the Pol I family is conserved in the RB69 structure, even though the detailed structures of the subdomains are quite divergent [7]. The exonuclease domains of Pol I and Pol II DNA polymerases are closely related in sequence and, not surprisingly, the structure of the exonuclease domain of RB69 resembles that of the Pol I type polymerases. Given the general similarity in the polymerase domains of the Pol I polymerases and RB69, the location of the exonuclease domain in RB69 was a surprise. In RB69 the 3′–5′ exonuclease domain is located above the fingers unit opposite the thumb subdomain, suggesting that the shuttling of DNA between the polymerization and proofreading sites must occur by a different mechanism in Pol II DNA polymerases [7].

The mechanism of the Pol I family DNA polymerases is now well understood [6,8,10,11,28]. The chemistry of nucleotide addition is mediated by two metal ions that are liganded by two aspartate residues. These are located in the palm subdomain, at the base of a deep cleft in the polymerase domain. High-resolution crystal structures of the Pol I type DNA polymerases of T7 bacteriophage (T7 Pol) and *Thermus aquaticus* (Taq Pol) complexed to primer–template DNA and incoming nucleotide have been determined, allowing the mechanisms of nucleotide incorporation and selectivity to be visualized [10,11,28]. Although corresponding structural information for the Pol II family DNA polymerases is lacking, similarities in general organization of the polymerase core as well as sequence conservation within crucial elements of the central palm subdomain suggest that general features of the recognition of DNA will be similar in Pol II polymerases.

The DNA polymerase from the archaebacterium *Sulfolobus solfataricus* strain Tok (D. Tok Pol) is a member of the Pol II family, and has both processable DNA polymerase and 3′–5′ exonuclease activities [12]. D. Tok Pol remains endonuclease DNA polymerase activity after incubation at 95°C for one hour (R.L., unpublished results). The sequence of D. Tok Pol is very closely related (> 70% identity) to that of other archaebacterial DNA polymerases, such as those from *Pyrococcus furiosus* [13] and *Thermococcus litoralis* [14]. D. Tok Pol is also related to eukaryotic DNA polymerases α, δ and ε (34% sequence identity over 90% residues of the DNA polymerase core for the human δ sequences) [1]. The archaebacterial genomes also contain genes coding for proteins with close homology to proliferating cell nuclear antigen (PCNA), the DNA polymerase clamp in eukaryotes, as well as subunits of the clamp-loader complex (RF-C (replication factor C)). It is likely that archaebacterial DNA polymerases achieve processivity by attachment to the ring-shaped PCNA ring, although direct evidence for such a mechanism is lacking.

We have determined the structure of D. Tok Pol at 2.6 Å resolution. D. Tok Pol shares less than 30% sequence identity with RB69 Pol, but the structures of the two enzymes resemble each other closely. The structure reported here has been determined in the absence of DNA. Nevertheless, the close structural correspondence between the active sites of Pol I and Pol II DNA polymerases allows inferences to be made about the mode of DNA recognition by D. Tok Pol. The very N-terminal region of D. Tok Pol possesses a domain (residues 5–132) that is closely related to archaeon or single-stranded RNA-binding domains (RBDs), also known as RNA-recognition modules (RRMs) [15]. The structure of the 3′–5′ proofreading exonuclease domain of D. Tok Pol is similar to those of the Pol I type polymerases. However, its location relative to the palm subdomain resembles the location seen in RB69 [7] rather than the Pol I type polymerases [8,10,11]. The structure of D. Tok Pol reported here provides further evidence that the mode of DNA-template recognition and the distant editing channel established for the Pol II family by the structure of RB69 Pol is valid for the entire Pol II family.

Results and discussion

Structure determination

Crystals of D. Tok Pol have been obtained from 2,5-methylpentanediol (MPD) (Native I) and polyethylene glycol (PEG) 400 (Native II). Both crystal forms are orthorhombic ($P2_12_12_1$, $a = 64.8$ Å, $b = 189.6$ Å, $c = 153.5$ Å for Native I and $a = 66.1$ Å, $b = 187.6$ Å, $c = 153.9$ Å for Native II). Experimental phases (Table 1) to 3.0 Å were obtained from four isomorphous heavy-atom derivatives, using Native II and the program SHARP [17]. Phases were improved by iterative cycles of real space density modification, consisting of solvent flipping and negative density truncation, using SOLOMON [18,19]. The resulting electron-density map allowed the chain to be traced unambiguously, with ready determination of sequence register. The model was refined to 2.6 Å against data for Native II (*R* value = 24.3%, R_{free} = 29.8%) and subsequently to 2.8 Å against data for Native I (*R* value = 25.3%, R_{free} = 29.9%), using CNS [20]. The model for Native II is somewhat more complete since the Materials and methods sections and it is used for most of the discussion. This model includes 746 residues from 1 to 756 in Native II. Amino acids 386–389 and 698–704 are not visible in our electron-density maps and are not included in the model.

General description of the structure

D. Tok Pol (Figure 1) is composed of a polymerase domain (residues 390–773) and an exonuclease domain (residues 133–385), as well as an N-terminal domain (residues 1–131) that is not found in Pol I type DNA polymerases [4]. The polymerase domain is further comprised of three smaller subdomains, termed the thumb (residues 667–756), palm (residues 390–442 and 582–690) and fingers (residues 448–499). The structures of the MPD and PEG400 crystal forms of D. Tok Pol are very similar

Table 1

Data collection, structure determination and refinement statistics.

	Resolution (Å)	Number of reflections (unique)	Completeness (%)	R_{sym}* (%)	R_{iso}† (%)	Sites	Phasing power‡	Figure of merit§
Native data								
Native 1	[illegible]	[illegible]	[illegible]	[illegible]				
Native 2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
MIRAS analysis								[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Refinement		Number of reflections (|F| > 2σ)		R_{cryst}¶/R_{free} (%)	Total number of atoms	Rmsd for bonds (Å)	Rmsd for angles (°)	Rmsd for B values (Å²)
Native 1	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Native 2	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

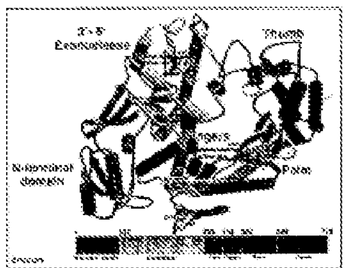
[illegible]

in terms of the individual subunits. The major difference between the two structures is a rotation of ~8–10° in the orientation of the exonuclease domain with respect to the thumb subdomain.

The domains of D. Tok Pol are arranged as an irregularly shaped flattened ring with a central cavity located near the polymerase active site. The mostly α-helical thumb subdomain forms one side of the active-site cleft and makes contacts with the exonuclease domain (Figure 3). The structures of the thumb domains of various polymerases are often unrelated in sequence. However, in all cases where structures are available the thumb domain is seen to fulfil an important role by forming contacts with duplex DNA as it exits the polymerase active site [4]. The D. Tok Pol structure has been determined in the absence of DNA, and a portion of the thumb subdomain that is likely to contact DNA (residues 665–675) is disordered. This is commonly observed for the corresponding regions of other polymerases in the absence of substrate [8,21–24]. In the DNA polymerases from bacteriophage T4 and RB69, the thumb subdomains also provide a C-terminal element that interacts with the processivity clamp [25,26]. In D. Tok Pol, the corresponding region (residues 757–773) is disordered.

The central region of the active-site cleft is occupied by the palm subdomain and includes residues important for substrate discrimination and the catalysis of the polymerase reaction. In D. Tok Pol, the palm is organized around three β strands (β16, β19, β20) flanked by an α helix (αQ) (Figures 1,2c,3e). It contains two disulfide bonds (Cys428–Cys442, Cys506–Cys509) that have not been previously observed in palm subdomains and which may be important for thermostability (Figure 3).

Figure 3



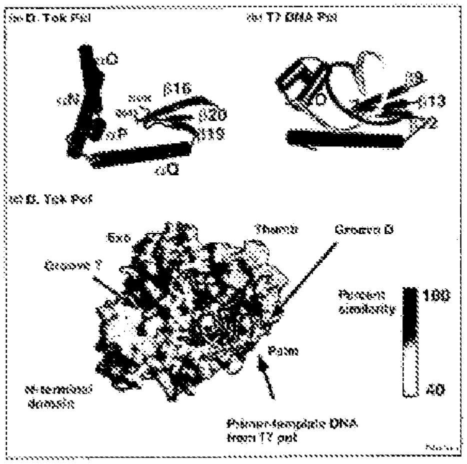
Structure of D. Tok Pol. The structure is represented by cylinders for helices, arrows for strands, and a thin ribbon for other secondary structure elements. [illegible] The active site of the polymerase is marked by the location of two aspartate residues D404 and D542. The two disulfide bonds are indicated. Regions of the polypeptide chain that could not be modeled in the palm subdomain because of disorder are indicated by dotted lines. The various domains and subdomains and their boundaries are indicated in the text.

USSN 10/089,027

Atty. Docket No. 2007651-0001

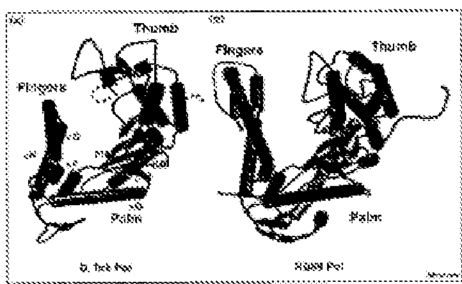
4567604v1

Figure 2



The central elements of the palm subdomains from polymerases belonging to the Pol I and Pol II families can be aligned closely (the root mean square deviation [rmsd] in Cα positions for strands β16, β19, β20 and helix αQ is in the range of 0.9–2.0 Å), indicating a potential conservation of function. There are two residues in the palm domains

Figure 3



USSN 10/089,027

Atty. Docket No. 2007651-0001

4567604v1

of Pol I polymerases that are crucial for enzymatic activity because they coordinate two metal ions [8,10,11,27]. The corresponding residues in D. Tok Pol are Asp408 and Asp542 (Figure 1). No metal ions are, however, visible in our electron-density maps.

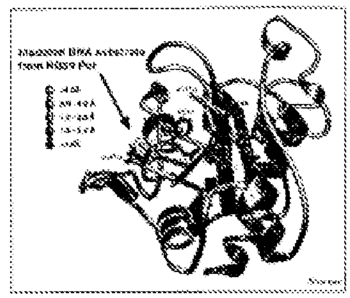
The fingers subdomain in D. Tok Pol consists of a set of antiparallel α helices (αN, αO, αP; Figure 2). These helices are shorter in length than the corresponding elements of RB69 Pol, and a helical segment that connects helices O and N in RB69 Pol is missing altogether (Figure 3). The fingers domain of D. Tok Pol is unrelated in overall structure to that of Pol I type polymerases (Figure 2). However, helix αP in D. Tok Pol is positioned similarly to helix O in Pol I polymerases (Figure 3), and is likely to play an analogous and crucial role in recognition of the incoming nucleotide [8–11,28].

The 3′–5′ exonuclease domain in D. Tok Pol is located opposite the thumb subdomain and above the fingers subdomain, as noted for RB69 Pol. It contains two metal ions (presumably Mg^{2+}) ligated to Asp141 and Glu143 (Figure 1). The position of this domain relative to the polymerase active site is distinct from the arrangement seen in Pol I type polymerases. The conservation between RB69 and D. Tok Pol of the location of the exonuclease domain suggests that this is a characteristic feature of Pol II type polymerases. The structure of the D. Tok Pol 3′–5′ exonuclease domain resembles those associated with other DNA polymerases [29,30]. The 3′–5′ exonuclease domains from the Pol I (E. coli, T. aquaticus, Bacillus subtilis, bacteriophage T7) or Pol II (RB69) polymerase families can be aligned onto each other closely (rmsd in Cα positions for strands β10, β11, β12, β14 and helices αE and αI is in the range of 1.0–2.8 Å). This alignment superimposes residues associated with substrate binding, catalysis and metal binding in a satisfactory manner (Figure 4) [9].

The arrangement of the N-terminal, exonuclease, and polymerase domains creates two deep grooves leading into and out of the polymerase active site. The D groove (for duplex-DNA binding, following the nomenclature of [7]) is located immediately below the thumb subdomain and includes a region of positive electrostatic potential. The T groove (for template-DNA binding) leads away from the active site in the opposite direction and is located below the fingers subdomain. A small channel (the editing channel) leads from the polymerase domain to the exonuclease active site (Figures 2c).

We have used the structure of T7 Pol bound to primer-template DNA to model DNA onto D. Tok Pol (Figure 2c). Superposition of the palm subdomains of the two polymerases shows that remarkably few bad contacts are formed between the DNA (from T7 Pol) and atoms in the D. Tok Pol model. The one region that does collide

Figure 4



Structural alignment of exonuclease domains. Structures of exonuclease domains from Kf, 1KFD, 1TTP, 1BGX, and 1NOY have been aligned by superimposing residues [illegible], which represent strands β8, β9, β10, β12, β13 and helix αE, αI. A color gradient is used to depict the average rmsd for the family of superimposed structures ranging from black (0–1.5 Å) to white (> 4.0 Å). Residues conserved amongst exonuclease sequences and implicated in catalysis are drawn in [illegible] representations. Two grey spheres represent two metal ions bound in the active site. The minor site is also indicated by a [illegible] derived from superposition of the exonuclease domain from the RB69 Pol structure.

with the DNA is the segment connecting the exonuclease and polymerase domains. This region (residues 377–385) is partially disordered in the D. Tok Pol structure, and is likely to reorganize upon binding DNA. This superposition allows five base pairs of DNA to be accommodated in the D. Tok Pol active site, with the formation of DNA-protein contacts. The formation of contacts with additional base pairs would require a change in the position of the thumb subdomain in the region of the D groove. A change in the conformation of the fingers subdomain (helices αO and αP) is also required to position residues Lys487 and Tyr491 (or Tyr494) of D. Tok Pol (Figure 2) for interaction with the incoming nucleotide, by analogy with the T7 Pol structure [11]. Finally, the superimposed primer-template DNA is well positioned so that the incoming template strand will probably reside in the T groove. Superposition of the DNA molecule derived from the structure of HIV-1 reverse transcriptase complexed to DNA [31] leads to similar conclusions.

Comparison between D. Tok Pol and RB69 Pol

Although the DNA polymerases from D. Tok Pol and bacteriophage RB69 share less than 20% primary sequence identity (Figure 3), their structures resemble each other

Figure 8

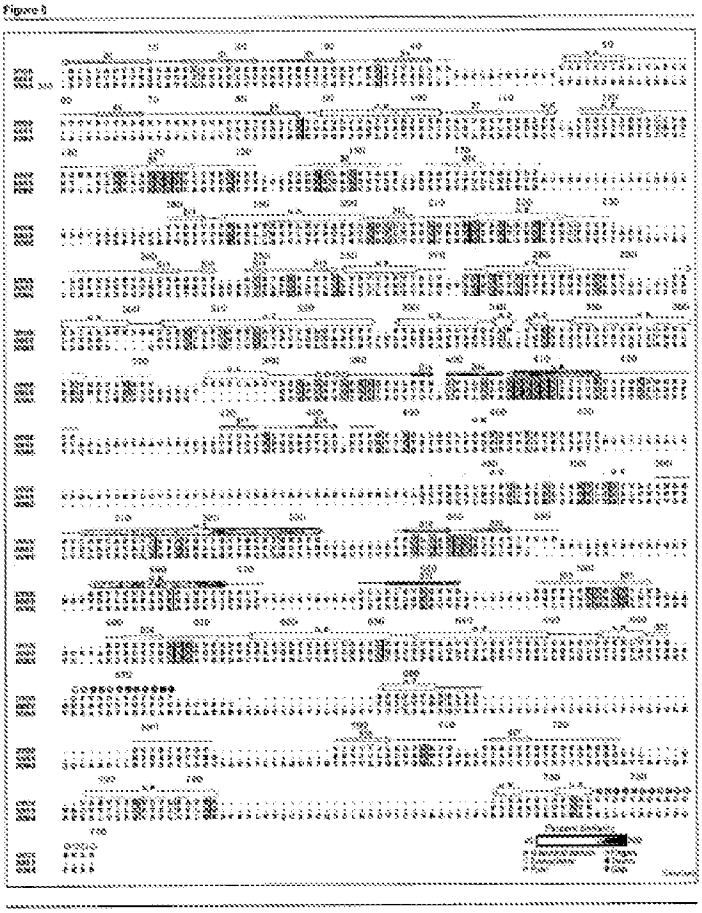


Figure 5

Structure based sequence alignment for D. Tok Pol (DTOK), RB69 Pol (RB69) and human Pol δ (HUMd). The RB69 sequence begins at residue 1??, as indicated by the number at the beginning of the sequence. The alignment is colored by sequence similarity [illegible] to 100%, green) calculated as described in Figure 3c. Shown here is a small subset of a larger set of sequences that were used to generate the alignment. [illegible]

closely (Figure 3). Not surprisingly, the regions of highest sequence similarity are concentrated in and around the exonuclease and polymerase active sites (Figures 3c,5). Despite the low overall sequence identity, the individual subdomains in the two structures superimpose well (the r.m.s. in Cα positions in the fingers, thumb and palm subdomains is in the range of 0.8 to 1.5 Å). Moreover, the overall arrangement of domains and subdomains with respect to each other is preserved in the two polymerases, strengthening the proposal that Pol B DNA polymerases share a common architecture (Figure 3).

One difference between the overall structures of D. Tok Pol and RB69 Pol concerns the orientation of the exonuclease domain with respect to the rest of the structure. When the two polymerases are superimposed on their respective palm subdomains it is seen that the exonuclease domain of RB69 is rotated inwards by ~8°, burying the active site in a solvent-inaccessible configuration [3]. In contrast, the exonuclease domain in D. Tok Pol has its active site essentially exposed to solvent. It is possible that conformational changes between open and closed configurations of the exonuclease domain are a part of the functional cycle of the protein, particularly as the two different forms of D. Tok Pol differ in the orientation of the exonuclease domain (not shown).

One interesting difference between D. Tok Pol and RB69 Pol is that the former is a thermostable DNA polymerase whereas the latter is not. Unfortunately, attempts to identify features in the D. Tok Pol structure that might be correlated with thermostability is complicated by the very low sequence similarity between the two enzymes. One feature that does stand out, however, is the increased formation of arrays of ionic interactions on the surface of D. Tok Pol when compared to that of RB69 Pol (Figure 6). The formation of networks of ionic interactions has been found to correlate with thermostability in other proteins [16,32,33].

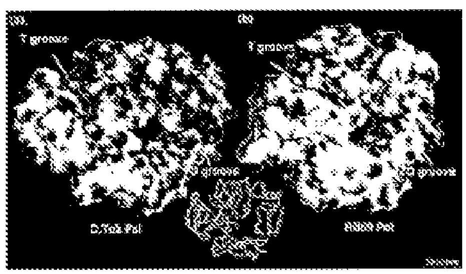
Generally, D. Tok Pol subdomains tend to be more compact, with smaller helices and shorter loops than are found in RB69 Pol, a feature that may be another important source of thermostability. For example, the palm subdomain displays close structural conservation of elements near the catalytic aspartate residues. However, helix αR in D. Tok Pol is much shorter than its counterpart in RB69 Pol, and a small subdomain in front of the palm subdomain is entirely missing in D. Tok Pol (Figures 3,5). Deletion of these elements is also seen in a representative set of archaebacterial DNA polymerases [13,34]. Likewise, the fingers subdomain is missing a large turn of from its tip in D. Tok Pol (Figures 3,5). However, the RB69 fingers extension most probably plays a T4 phage-specific role, as it is also missing from our alignments of archaebacterial DNA polymerases and eukaryotic polymerases δ (Figure 5).

The N-terminal domain resembles RNA-binding domains

The N-terminal domain of D. Tok Pol has no corresponding element in Pol I type polymerases. Analysis of the

Figure 6

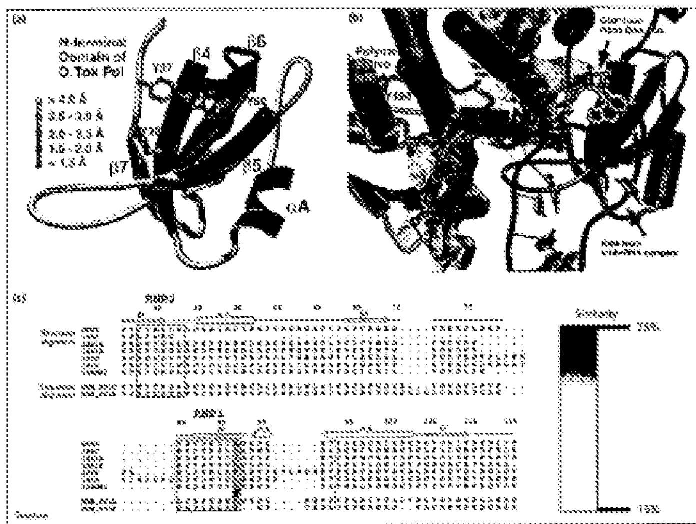
Comparison of surface charges in D. Tok Pol and RB69 Pol. Accessible surface representations of (a) D. Tok Pol and (b) RB69 Pol in the same orientation after superposition [illegible]



USSN 10/089,027

Atty. Docket No. 2007651-0001

4567604v1


N-terminal
Domain of
D. Tok Pol
β4
β6
β7
β5
αA
Similarity
75%
15%

and bind to single-stranded RNA. Two conserved sequence motifs, referred to as RNP1 (ribonucleoprotein 1) and RNP2, provide aromatic and charged residues that are important for RNA recognition [35] (Figure 7).

The N-terminal domain of D. Tok Pol can be superimposed closely onto the core secondary structural elements of RBDs from the U1A spliceosomal protein [35], ribosomal protein S6 [36], the heterogeneous ribonucleoprotein (hnRNP) proteins (two RBD domains) [37,38] and the anticodon-binding domain from T. thermophilus phenylalanyl-tRNA synthetase [39]. The rmsds in Cα positions for these superpositions are in the range of 1.5–2.0 Å (Figure 7a). Differences between the structures of the loops in the N-terminal domain of D. Tok Pol and those of the RNA-binding domains are within the range of structural variation seen in the various RNA-binding domains.

There is no evidence at present to suggest that the N-terminal domain of D. Tok Pol binds RNA. However, comparison with the structures of RNA complexes of RNA-binding domains shows that the N-terminal domain might in fact be a functional RNA-binding domain (Figure 7). In particular, three aromatic residues in the N-terminal domain (Tyr37, Tyr39 and Tyr86) could interact with RNA bases in a manner similar to that seen in crystal structures of RNA bound to RNA-binding domains [35] (Figure 7). Interestingly, these residues are located near the position of a guanosine triphosphate molecule that is found bound to the N-terminal domain of RB69 Pol [1] (Figure 7b). The DNA polymerases from bacteriophage T4 and its distant relative bacteriophage RB69 bind specifically to the ribosome-binding site of their own mRNA (messenger RNA), repressing its translation [40–42]. The N-terminal domains of T4 Pol and RB69 Pol are smaller than that of D. Tok Pol. In the RB69 Pol structure, the N-terminal domain seems to form an 'incomplete' RNA-binding domain (Figure 7c).

There is no significant overall sequence similarity between the N-terminal domain of D. Tok Pol and RNA-binding domains, which is why the presence of this fold was not recognized previously (Figure 7c). Comparison of the sequences of other archaebacterial DNA polymerases and human polymerases δ and ε suggests that a corresponding structural element is likely to be found in these polymerases as well (Figure 7c). The sequence alignment in this region is unambiguous for the archaebacterial DNA polymerases. For eukaryotic polymerases the alignment is less certain, but it seems to conserve the essential aromatic character of the RNP motifs (Figure 7c). Confirmation of the presence of these domains along with their ability to bind RNA, and their precise role in eukaryotic DNA synthesis awaits future structural and functional studies.

Biological implications

The structure of the DNA polymerase from the archaebacterium *Desulfurococcus* strain Tok, D. Tok Pol, reveals a strong similarity to the DNA polymerase from bacteriophage RB69. It also reveals the presence of an N-terminal domain that has structural similarity to RNA-binding domains from the U1A spliceosomal protein, ribosomal protein S6, the hnRNP proteins and the anticodon-binding domain from *T. thermophilus* phenylalanyl-tRNA synthetase. Although the structure in the immediate vicinity of the central catalytic region of the polymerase domain closely resembles that of Pol I type DNA polymerases, the overall architecture of D. Tok Pol and the placement of the exonuclease domain is strikingly different. The similarity between D. Tok Pol and RB69 Pol suggests that these two structures are representative of a common Pol II polymerase fold. Members of this family carry out chromosomal DNA replication in eukaryotes, including humans, and yet there is no structural information available for any eukaryotic member of this family. While this manuscript was being prepared, the structure of another archaebacterial DNA polymerase, that from the organism *Thermococcus gorgonarius* has been reported [58]. The D. Tok Pol structure reported here, along with the RB69 Pol structure and the structure of the *Thermococcus gorgonarius* DNA polymerase, should now make it possible to generate reliable structural models for eukaryotic DNA polymerases.

Materials and methods

Protein expression and purification

[illegible]

Crystallization, cryocrystallization, and heavy-metal derivatization

[illegible]

USSN 10/089,027

Atty. Docket No. 2007651-0001

4567604v1

Data collection and phase determination

Model building and refinement

Figure preparation

Accession numbers

Acknowledgements

References

21. Kornberg, R. & Baker, T.A. (1992) Structure of T4 DNA polymerase complexed to the nonincorporated inhibitor T3 synaptic. *FASEB J.* 12, 1409–1413.
22. Kunkel, T.A., Bebenek, K. & Wang, B.C. (1990) The Klenow fragment, lacking its fourth subdomain, of T7 DNA polymerase is processed by the exonuclease of a chimeric T7/E. coli DNA polymerase. *J. Biol. Chem.* 265, 6–12.
23. Kim, Y., Eom, S.H., Wang, J., Lee, D.-S., Suh, S.W. & Steitz, T.A. (1995) Crystal structure of *Thermus aquaticus* DNA Polymerase. *Nature* 376, 612–616.
24. Jacobo-Molina, A., et al. & Arnold, E. (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc. Natl Acad. Sci. USA* 90, 6320–6324.
25. Greenwich, J.A., Lee, T.C., Sawai, E.T., Jones, C. & Konigsberg, W.H. (1993) Residues in the carboxy terminus of T4 DNA polymerase are important determinants for interaction with the polymerase accessory proteins. *Biochemistry* 32, 10714–10721.
26. Bonner, R.L., Sawaya, M. & Beese, L. (1996) The carboxy terminus of the bacteriophage T4 DNA polymerase is important for holoenzyme complex formation. *Proc. Natl Acad. Sci. USA* 93, 12222–12227.
27. Steitz, T.A., Smerdon, S.J., Jager, J. & Joyce, C.M. (1994) A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* 266, 2022–2025.
28. Kunz, B.A., Kohalmi, S.E., Kunkel, T.A., Mosbaugh, D.W., Nishimura, R.N., Taylor, J.S. & Bonner, J.J. (1989) Mutational specificity of yeast DNA polymerase delta in a recombination system. *Science* 244, 1081–1083.
29. Minnick, D.T., Bebenek, K., Osheroff, W.P., Turchi, J.J., Kunkel, T.A. & DeVoe, H. (1999) Side chains that contribute to the fidelity of DNA polymerase zeta. *J. Biol. Chem.* 274, 3067–3075.
30. Kiefer, J.R., et al. & Beese, L.S. (1997) Crystal structure of a thermostable Bacillus DNA polymerase I large fragment at 2.1 A resolution. *Structure* 5, 95–108.
31. Wang, M., Gregra, R., Yardley, G.J. & Morrison, S.D. (1999) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* 285, 1068–1074.
32. Yu, R.C., et al. & Connolly, B. (1996) The structure of *Pyrococcus furiosus* ligase-like thermophilic deoxyribonucleic acid ligase for forced changes in nucleotide cofactor specificity of substrate recognition. *Structure* 4, 1463–1480.
33. Hopfner, K., Eichinger, A., Sauer, R., Steitz, T. & Aertsen, A. (1999) Structures of endo-exonuclease VIII from *Escherichia coli* in complex with DNA: catalysis and substrate selection of a base-excision repair enzyme lacking a metal cofactor. *J. Mol. Biol.* 299, 699–711.
34. Wilson, D. & Sancar, A. (1995) Protein motifs and homologs: sequence and structure alignments. *Nucleic Acids Res.* 23, 3465–3471.
35. Delagoutte, E., Jr., G., Eyrisch, S., Yan, C.-H. & Baase, W. (1999) Crystal structure of T7 RNA polymerase of the DNA-binding domain of the DNA polymerase protein complexed with an RNA hairpin. *Nature* 399, 699–708.
36. Lesburg, M., et al. & Wiechens, J.A. (1999) Crystal structure of the transcription protein S6 from *Thermus thermophilus*. *EMBO J.* 18, 1238–1250.
37. Sharma, V., Knaggs, M., Rice, L.M., Grishaev, A.B. & Steitz, T.A. (1999) Crystal structure of the two DNA binding domains of human hnRNP A1 at 1.75 A resolution. *Nat. Struct. Biol.* 6, 715–722.
38. Xia, S.M., Jaiman, L., Chang, J., Mayeda, A. & Krainer, A.R. (1999) Crystal structure of human SF2, the domain of hnRNP A1 that recognizes 5' splice site RNA recognition motifs. *Structure* 8, 859–870.
39. Eichinger, C., et al. & Stark, M. (1997) The crystal structure of glycogenin: RNA synthesis from sequence homologies complexed with reagent. *EMBO Rep. Structure* 5, 59–68.
40. Wang, C.C., Hsieh, K. & Keown, W.D. (1992) Excision of RNA priming oligonucleotides by T4 DNA polymerase. *J. Biol. Chem.* 267, 23799–23806.
41. Goodman, M.F., Bebenek, K., Petruska, J. & Creighton, S. (1988) Exonucleolytic proofreading by calf thymus DNA polymerase delta: modulation of fidelity with dNTP pool composition and the editing efficiencies of 24 single base pair mismatches. *J. Biol. Chem.* 263, 20178–20184.
42. Parker, A.P. & Bonner, J.D. (1994) Binding specificity of T4 DNA polymerase to RNA. *J. Biol. Chem.* 269, 19858–19862.
43. Charbonnier, F. & Mergy, W. (1999) Processing of 4-day-detected data collected in successive scans. *Methods Enzymol.* 276, 307–326.
44. Rossi, R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* 42, 340–349.
45. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110–119.
46. Rice, L.M., Shamoo, Y. & Brunger, A.T. (1998) Phase improvement by multiple molecular envelope averaging. *J. Appl. Crystallogr.* 31, 798–805.
47. Brunger, A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475.
48. Collaborative, C.C.P., MacArthur, Blow, Bricogne, G.N., S Charnock, Dodson, E.J., Evans, P.R., Hudson, K.A., Jones, E.Y., Kirby, R., McGrath, J., Morris, R.J., Pearce, N.M., Potterton, E.A., Potterton, L., Roberts, S.M., Taylor, J.C., Wilkinson, A.J., Wilson, K.S., Wolf, W. & Woo, E.J. (1994) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.
49. Carroll, M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* 15, 128–130.
50. Nicholls, A., Sharp, K. & Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281–296.
51. Jones, S. (1997) DSSP2PS 2.0. *J. Appl. Crystallogr.* 24, 956–961.
52. Gouet, P. (1999) ESPRIPT: Analysis of multiple sequence alignments. *Protein Eng.* 6, 27–32.
53. Nagai, K., Oubridge, C., Jessen, T.H., Li, J. & Evans, P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* 348, 515–520.
54. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
55. Hendrickson, W. & Konnert, J.H. (1980) Biomolecular structure and function. *Proc. Natl Acad. Sci. USA* 88, 9309–9313.
56. Fletcher, R.P., et al. & Ferguson, D. (1998) Crystal structure of a deoxyinosine-specific endonuclease. *Nature* 394, 381–386.

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 3600–3605, March 1999
Biochemistry

Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*

(x-ray structure/disulfide bonds/replication/Archaea/exonuclease)

KARL-PETER HOPFNER*†‡, ANDREAS EICHINGER*, RICHARD A. ENGH*§, FRANK LAUE§, WALTRAUD ANKENBAUER§, ROBERT HUBER*, AND BERNHARD ANGERER§

*Abteilung Strukturforschung, Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany, and §Roche Diagnostics, D-82377 Penzberg, Germany

Contributed by Robert Huber, January 22, 1999

ABSTRACT Most known archaeal DNA polymerases belong to the type B family, which also includes the DNA replication polymerases of eukaryotes, that maintain high fidelity at extreme conditions. We describe here the 2.5 Å resolution crystal structure of a DNA polymerase from the Archaea *Thermococcus gorgonarius* and identify structural features of the fold and the active site that are likely responsible for its thermostable function. Comparison with the mesophilic B type DNA polymerase gp43 of the bacteriophage RB69 highlights thermophilic adaptations, which include the presence of two disulfide bonds and an enhanced electrostatic complementarity at the DNA–protein interface. In contrast to gp43, several loops in the exonuclease and thumb domains are more closely packed; this apparently blocks polymer binding to the exonuclease active site. A physiological role of this "closed" conformation is unknown but may represent a polymerase mode, in contrast to an editing mode with an open exonuclease site. This archaeal B DNA polymerase structure provides a starting point for structure-based design of polymerases or ligands with applications in biotechnology and the development of antiviral or anticancer agents.

Propagation of cells requires faithful DNA replication. This is performed in vivo by DNA polymerases (pols), which attach the appropriate dNTP to the nascent DNA primer strand in contact to a paired template. Different families of pols are involved in different DNA polymerization processes including not only DNA replication (1, 2) but also repair and recombination (3, 4), a heterogeneity also reflected by varying polypeptide structures and/or subunit compositions (5, 6). Some pols exhibit 3′ → 5′ exonuclease activity (editing activity) and/or 5′ → 3′ "structure-specific endonuclease" activity, often located in separate structural elements of the same polypeptide chain (6–8).

Crystal structures are available for most known polymerase families, including the A family DNA polymerases (9–14), pol β (15–17), HIV reverse transcriptase (18–20), and recently, the B family pol gp43 from bacteriophage RB69 (21). All share a functional polymerase structure, which resembles a right hand built by the palm, fingers and thumb domains (see ref. 7 for review). Although the fingers and thumb domains are highly diverse among the different families, the palm domains, which contain the conserved catalytic aspartate residues, show a similar topology among all families except pol β. The polymerase nucleotidyl transfer was studied in detail for the A family polymerases, HIV reverse transcriptase, and pol β, and was shown to involve two metal ions (summarized in ref. 7).

Considerably less is known for the family of type B pols, which are replicative enzymes in eukaryotes and most likely also Archaea (22, 23). The structure of gp43 from bacteriophage RB69 (21) provided an excellent first insight into this family. In addition to the three polymerase domains, gp43 contains an 3′ → 5′ exonuclease domain and an N-terminal domain. The exonuclease and palm domains share the topology and active site of A family enzymes, implying similar metal-assisted mechanisms for polymerase and exonuclease activities (21). The thumb and finger domains are apparently unrelated to the other polymerase families. The function of the N-terminal domain remains unknown, but may help assemble the multicomponent replication apparatus (21).

Much is known about the replication of phages (24–26), viruses (1, 27), Prokaryota (28) and Eukaryota (1, 3, 29, 30), which in general involves pols that solve primosome, helicases, SSBs/RPA, sliding clamps, and other factors (31). Considerably less is known for archaeal replication, where mostly B type polymerases, similar to eukaryotic replication enzymes pol α and δ, have been identified (6, 22, 23, 32–34). This relative ignorance is surprising, because such crystal biotechnological applications as cloning and PCR require the thermostability and fidelity typical of archaeal polymerases (6). Thus, in addition to satisfying basic research interests, structural information could assist, for example, the engineering of variant enzymes with tailored nucleotide incorporation rates or the design of antiviral and anticancer polymerase inhibitors. For these reasons, we have determined the structure of a DNA polymerase from *Thermococcus gorgonarius* (*Tgo*), an extremely thermophilic sulfur metabolizing archaeon isolated from a geothermal vent in New Zealand (35). This enzyme possesses pol and a 3′ → 5′ exonuclease activity, which together ensure thermostable replication with high fidelity (error rate: $5.5–2.3 \times 10^{-6}$; see ref. 36). The 2.5 Å structure shows a topological similarity to gp43 and gives insight to the conserved biology of archaeal DNA polymerases, including the identification of several mechanisms for thermophilic adaptation.

MATERIALS AND METHODS

Materials. All materials were of the highest grade commercially available.

Bacterial Strains. *Escherichia coli* LE392 containing pUC88520 was used as described (36). *E. coli* BL21 (DE3) (host motif) was a generous gift of Rodrigo Bohm (Max-Planck-Institut).

Expression Vectors. pET21 was obtained from Novagen Molecular Biochemicals.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: *Tgo*, *Thermococcus gorgonarius*; pol, DNA polymerase; Exo, exonuclease. The atomic coordinates have been deposited in the Brookhaven Data Bank, Biology Department, Brookhaven National Laboratory, Upton, NY 11973 (PDB ID code 1TGO).

†Present address: The Scripps Research Institute, La Jolla, CA 92037.
‡To whom reprint requests should be addressed. e-mail: hopfner@scripps.edu.

Table 1. Data collection and isomorphous replacement

Derivative	Resolution limit, Å	Total observations	Unique observations	Completeness, %	R_{sym}, %	R_{iso}, %	Phasing power
Native	3.0	136,978	23,229	91.9	7.0		
U	3.2	25,961	16,155	66.5	11.5	58.5	0.75
PT1	3.7	65,386	13,882	96.2	12.0	22.6	1.58
PT2	4.0	23,820	6,293	66.0	15.5	12.8	1.77
PT3	3.4	85,326	14,152	93.8	10.5	17.5	1.56
PT4	2.6	387,923	25,485	90.6	5.7	35.3	0.62
PT5	2.7	350,665	24,543	88.7	6.1	36.2	0.53
PTU	3.5	79,620	12,437	82.4	9.6	26.7	1.34
PB	3.5	78,150	11,336	91.9	11.6	32.8	0.41
PBPT	3.4	66,557	14,129	91.6	8.0	39.6	1.04
OS	2.8	658,168	24,036	98.2	5.5	38.8	1.85

Overall figure of merit (25.0–3.0 Å): 0.73

Heavy-atom derivatives were prepared by soaking the crystals in low-salt buffer containing the heavy atom as follows: U, 0.5 mM uranyl acetate for 2 h; PT1, 5 mM K_2PtCl_4 for 1 d; PT2, 5 mM K_2PtCl_4 for 2 d; PT3, saturated cis-dichlorodiaminoplatinum(II) for 3 d; PT4, 5 mM K_2PtCl_4 for 7 d; PT5, saturated cis-dichlorodipyridine-Pt(II) for 3 d; PTU, PT1 + U; PB, saturated diaminophenyl-Pb(NO_3)$_2$ for 2 d; PBPT, PB + PT5; OS, saturated K_2OsO_4 for 7 d. NAT1, PT1, PT2, PT3, PTU, PB and PBPT were collected with a Mar imaging plate and U was collected with a Bruker AXS area detector on a Rigaku rotating anode source. All other data sets were collected with a Mar charge-coupled device at beamline BW6 at DESY, Hamburg.

Cloning, Expression, and Protein Purification. The gene for the 89.8-kDa DNA-dependent pol (*Tgo* pol) was cloned from *Tgo* (Deutsche Sammlung von Mikroorganismen no. 8976) and expressed in *E. coli* LE392pUBS520 pSite2Tgo (Deutsche Sammlung von Mikroorganismen no. 11328) as described (35). *Tgo* pol was purified essentially as described (36) with the substitution of the TSK butyl Toyopearl column by Blue-Trisacryl M (Sepra) and with an additional concentration step on Poros 50 HQ anion exchange medium (Boehringer Mannheim Biochemicals). Active fractions were combined, concentrated to 30 mg/ml, and transferred to 20 mM sodium phosphate, pH 8.2/10 mM 2-mercaptoethanol/500 mM NaCl.

The gene for a selenomethionine-containing variant of *Tgo* pol (Se-*Tgo* pol) was expressed in *E. coli* B834 (DE3) (hsd metB) using a published protocol (37). Se-*Tgo* pol was purified by using the wild-type protocol.

Crystallization. Crystals of purified *Tgo* pol (or Se-*Tgo* pol) were grown by using sitting-drop vapor-diffusion techniques at 4°C with high-salt conditions (2:2 µl protein:reservoir—100 mM Tris, pH 8.7/2.0M ammonium sulfate) and diffracted to 3.0 Å (in-house) and to 2.5 Å [beamline BW6 at Deutsches Elektronen Synchrotron (DESY), Hamburg]. Low-salt conditions (100 mM Tris, pH 7.0/200 mM ammonium sulfate/30% PEG 400) yielded only poorly diffracting crystals but allowed soaks (including heavy atoms) with some cell constant modulations (a, b, c = 63.6, 109.0, 160.5) but minimal loss of resolution.

Data Collection and Processing. Data were collected with a MAR imaging plate or a Bruker AXS X1000 mounted on a Rigaku rotating anode source, or with a MAR imaging plate or a MAR CCD (charge-coupled device) at beamline BW6 at DESY, Hamburg. The data were processed with SAINT (Bruker AXS), MOSFLM (Mar CCD; ref. 38), or DENZO (MAR imaging plate; ref. 39), scaled with SCALA (40) or SCALEPACK (39), and reduced with TRUNCATE (40).

Structure Determination. The structure was solved by multiple isomorphous replacement and anomalous scattering (MIRAS) by using data from crystals transferred to low-salt conditions (Table 1). Crystallographic calculations were done with programs from the CCP4 suite (40). Heavy atom positions of major sites were located in difference Patterson maps and were refined with MLPHARE (40) to calculate protein phase angles to 3.5 Å resolution. A partial polyalanine model was built into interpretable portions of secondary structural elements of the MIRAS map by using MAIN (41). The quality of the electron density was improved by phase combination of the partial model with the experimental phases by using SIGMAA (40), and several cycles of solvent flattening to 3.0 Å by using SOLOMON (40). At this stage, no interpretable density was found for a significant portion of the molecule, comprising residues 147–158, 225–306, 653–728, and 752–773.

Model Building and Refinement. The partial model (R factor 35%) was used to phase the 2.5 Å resolution data of the Se-*Tgo* pol (high-salt conditions). The model was oriented with AMORE (40). The correlation coefficient of 22.0% and the R factor of 50.3% showed divergence of the high- and low-salt structures. After bulk solvent correction, anisotropic B factor correction and rigid-body minimization (treating five domains independently), the partial model was iteratively refined and extended with simulated annealing, Powell minimization, restrained individual B factor refinement with CNS (42), and manual model building with MAIN (41) by using data from 25.0–2.5 Å resolution (Table 2, Fig. 1).

RESULTS AND DISCUSSION

Structure of *Tgo* pol. *Tgo* pol is a ring-shaped molecule with dimensions 50 Å × 80 Å × 150 Å. The single polypeptide chain of 773 aa is folded into five distinct structural domains (Fig. 2): the N-terminal domain (residues 1–130), the 3′ → 5′ exonuclease domain (131–326), the palm (369–449 and 500–585), fingers (450–499), and thumb (586–773) domains of the polymerase unit, and a helical interdomain insertion (327–368).

Table 2. Crystallographic refinement, high-salt form

Space group	$P2_12_12_1$
Cell dimensions, Å	a = 58.1, b = 108.2, c = 156.2
Observations, 25–2.5 Å	
Total	232,448
Unique	30,431
Completeness, %	
Total	93.1
Last shell	86.6
R_{sym}, %	
Total	3.2
Last shell	30.2
R factor (R_{free}), %	20.8 (27.5)
rms deviation in bond lengths, Å	0.008
rms deviation in bond angles, °	1.5
No. of nonhydrogen atoms	
Protein	6,358
Water	339

FIG. 1. Stereoscopic representation of the electron-density map. The $2|F_o| - |F_c|$ electron density (contoured at 1.0 σ) at the polymerase active site is well defined for the refined model (stick representation).

between the exonuclease and palm domains. The polymerase unit forms the DNA-binding crevice, reminiscent of a right hand, which is the identifying characteristic of pols. Gp43 from bacteriophage RB69 also shows this overall domain topology (21).

Three clefts extend radially from the polymerase active site at the center of the ring; two of them in opposite directions, forming a large cleft across the molecule, and one perpendicular to these. Based on active-site homology to pol A family enzymes, the two opposite clefts probably bind duplex DNA (cleft D, according to ref. 21) and single-stranded template DNA (cleft T), respectively. The perpendicular (editing) cleft links the polymerase active site and the exonuclease active site and binds the primer strand in editing mode (21).

The exonuclease domain is structurally equivalent to the 3′ → 5′ exonuclease domain of pol A family (45). Like gp43, however, it is bound at the opposite side of the polymerase unit by noncovalent contacts to the thumb domain at the editing cleft, on one side, and by covalent and noncovalent contacts to the N-terminal palm domain and the 32 residue interdomain helix, on the other side. This latter segment is located at the base of cleft T, which is additionally bounded by the exonuclease, N-terminal, and palm domains.

The topology of the palm domain is conserved among polymerase families (5), with two long helices (Q and R)

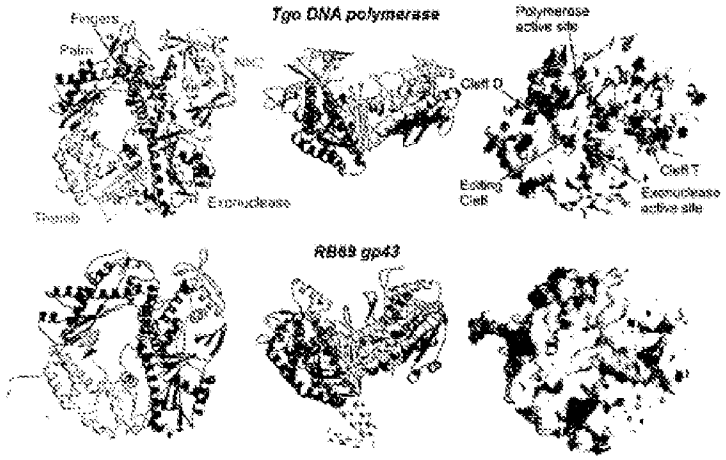


FIG. 2. Structure of Tgo pol and comparison with gp43 from bacteriophage RB69. (*Left* and *middle*) Ribbon representation of Tgo DNA polymerase (*Middle*) in two orthogonal views with secondary structure elements. The molecule is composed of five domains. N-terminal domain (yellow), 3′ → 5′ exonuclease (red), palm (light and dark magenta) (segment the N-terminal and C-terminal segment, respectively (see text)), fingers (blue), and thumb (green), which are arranged to form a ring. An interdomain helical segment between the exonuclease domain and the palm is orange. The conserved carboxylates in the active site and the two disulfide bridges are shown as magenta and yellow ball-and-stick, respectively. Tgo pol has the same overall architecture and domain topology (here gp43 of RB69 (*Lower*). The N-terminal insertion in the fingers of gp43 is gray. (*Right*) Comparison of molecular surfaces of Tgo pol (*Upper*) and RB69 gp43 (*Lower*). Red and blue denote negative and positive electrostatic surface potential, respectively. In contrast to gp43, Tgo pol has a strongly enhanced positive potential at the putative DNA-binding cleft.

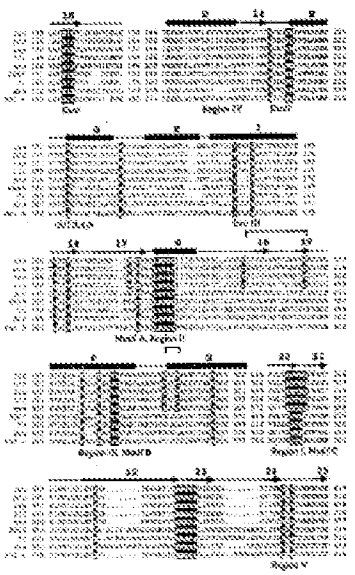


FIG. 3. Sequence alignment of B family DNA polymerases. The alignment has been adapted from ref. 21 to highlight specific contacts from the class of archaeal pols. The secondary structure of Tgo pol is indicated on top of the alignment with helices (bars), strands (arrows) and loops (lines) colored according to domains with the same color code as Fig. 2B. Strictly conserved residues of type B polymerases are red, and additional conserved residues are yellow. Uniquely conserved residues of archaeal type B sequences—as discussed in the text—are green. Disulfide bonds are shown by a bar on top of the alignment. Abbreviations: pol, *Thermococcus gorgonarius* pol; gla, *Pyrococcus furiosus* pol; vgo, *Thermococcus* sp. pol; 9°, *Thermococcus litoralis* pol; ones, *Methanococcus voltae* pol; RB69, bacteriophage RB69 pol; T4, bacteriophage T4 pol; Rlic, *E. coli* pol II; pol B, bottom pol I.

packed against the five-stranded antiparallel β-sheet that contains the three conserved aspartate residues involved in nucleotidyl transfer. The fingers converge from the palm domain as an α-helix-rich insertion. Six 30 residues are folded into two antiparallel coiled α-helices of approximately equal size; helix P contains the conserved $Ku_3Na_2Y(Ex_3)G$ motif of B type polymerases and is related to the O helix of A type enzymes (see below). The ≈50 residue insertion between helix O and P in RB69 and T4 gp43 is missing in Tgo pol, where both helices and 4 residue linker are much shorter than their equivalents in gp43. The shorter fingers of Tgo pol presumably reflect the typical structure of the mesophilic/thermophilic B type fingers (pol α, pol δ, and *E. coli* pol II). The thumb domain topology, similar to that of gp89, is unrelated to other polymerase types. However, like the thumb of A type enzymes, a bundle of α-helices at its base protrudes from the active site β-sheet platform at the active site, the thumb contains a 75-residue subdomain (665–729), which forms the exonuclease domain and contributes to the cutting channel, explaining why mutations in the exonuclease domain of B-type polymerases affect the polymerase activity and vice versa (44, 45).

Weakly defined density across the base of the thumb domain was modeled as the C-terminal 6 residues with a polyalanine chain. The C terminus then does not protrude from the core molecule as in the RB69 polymerase (21). Because the C terminus of the T4 pol are involved in sliding-clamp binding (46), it is likely, however, that these residues become ordered on any similar holoenzyme formation.

Sequence Alignment of Archaeal DNA Polymerases. The structure of Tgo pol allows the generation of a structure based sequence alignment of the archaeal subfamily of type B DNA polymerases, the location of conserved and unique residues, and the comparison with other type B DNA polymerases (Fig. 3).

Polymerase Active Site. The polymerase active site is formed by the central β-sheet (strands 16, 17, 20, 21, and 22) and helix N of the palm domain and helix P located in the fingers and is highly conserved among B family polymerases (Fig. 4). Three carboxylates required for nucleotidyl transfer in B family polymerases, two of which coordinate two metal ions (34) are superimposable conserved among A family enzymes, B family enzymes, and reverse transcriptase (21). Superposition of Tgo pol and T7 replication complex (14) places the dNTP near the proposed nucleotide-binding site in helix P, the $K_{48}I_{49}N_{50}Y(Ex_3)S$ motif (Fig. 3) and suggests interactions of the carboxylates with metals and the phosphate tail of the bound dNTP (Fig. 4). Superimposition of the strictly conserved Lys-487 allows it to mimic the Lys-522–phosphate salt interaction in T7. Tyr-488 ($K_{48}I_{49}N_{50}Y_{488}I_{Ex_3}S$) and Tyr-409 ($RLY_{409}PSIE$) form the floor of the nucleotide binding site.

The active site of B family polymerases contains a DTDS motif, which, however is DTDG in the archaeal subfamily. In Tgo pol, the relatively conserved Tyr-402 from the adjacent strand provides an aberrant alcohol group, it is probably appropriate for metal coordination or binding of the 3′ end of the primer. The orientation of Tyr-402 is stabilized by an aromatic cluster that also includes Phe-543 and Tyr 538. Archaeal *Methanococcus voltae* and *Thermococcus* sp. pol 9°N (see Fig. 1) have Tyr at position 538—Phe in Tgo—rather than at 402, but might also supply an alcohol group. The displacement of a functional alcohol from serine in DTDS to Tyr-402 or Tyr 538 might stabilize its orientation as an adaptation for thermostability.

The conserved cluster of acidic amino acids (D578, E580) forms an unexpected metal-binding site for Mn^{2+} and Zn^{2+} (Fig. 4) its proximity to Asp-404 and to the expected location of the dNTP γ-phosphate suggests a supporting role to nucleotide binding and/or catalysis.

3′ → 5′ Exonuclease Active Site. Tgo pol is characterized by a strong 3′ → 5′ exonuclease activity, unlike eukaryotic B type polymerases (unpublished results). The exonuclease active site is located at the interface between the exonuclease domain and the tip of the thumb (Fig. 5). All residues required for catalysis are located in the exonuclease domain, which, at least for T4 gp43, retains activity when dissociated from the polymerase (47). However, the thumb domain, with, for example, RB69 gp43's Phe-122–Tyr-130 intercalation, partially controls the binding geometry of single-strand DNA (21, 45).

The exonuclease structures of Tgo and gp43 DNA polymerases are similar at the editing site but differ considerably at the exonuclease–thumb interface. Strand 19 contains the metal-binding D381E motif and readily superimposes with the equivalent strand from gp43, allowing modeling of a single-strand DNA segment into the exonuclease site based on the RB69 gp43–$p(dT)_6$ complex (21). The conserved residues Asp-381 and Glu-383 in the Exo I motif, Tyr-388, Asn-389, Phe-458, and Asp-215 in Exo II, and Tyr-341 and Asp-343 in Exo III are in approximate DNA-binding conformations (Fig.

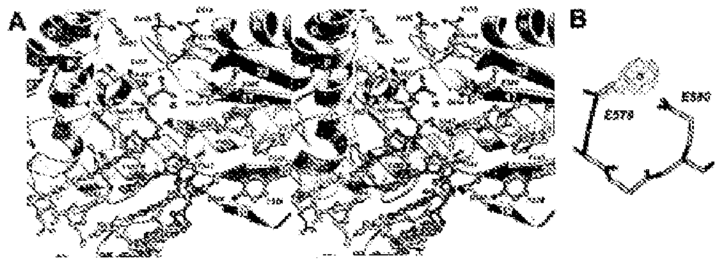


FIG. 4. Polymerase active site. (A) Stereoribbon representation (using color code as in Fig. 2) with modeled DNA. Active-site residues are shown as ball-and-stick representations with carbon (green), nitrogen (blue), and oxygen (red) atoms. The DNA template (light brown), primer (light brown), and dNTP (orange) complex has been taken from the coordinates of T7 replication complex (32) by superimposing D404, D542, and adjacent residues with corresponding residues in T7 pol (D475 and D654). Phosphorus atoms are yellow. The two metals of the T7 replication complex are shown as magenta spheres. (B) Experimentally observed metal-binding site for Mn^{2+} ($F_o - F_c$ density contoured at 5σ) and Zn^{2+} in the "low salt" crystal form. The carboxylates E578 and E580 are conserved in type B polymerases (Fig. 2).

54). However, the editing cleft is constricted by a displacement of the tip of the thumb toward the exonuclease domain to prohibit single-strand binding (Fig. 5B). This shift is correlated with a large change in the loop between strands 18 and 15. In RB69 gp43 (and likewise in T4 gp43), this loop forms a lid over the 3′ base and contains Phe-123, which intercalates between the first two bases. In *Tgo* pol, the loop is turned outward, away from the thumb, and Phe-152 (the equivalent of gp43's Phe-123) attaches to Phe-214 10 Å away from intercalation site. This shift allows the tip of the thumb to move into the editing channel and to block the exonuclease site.

Are There Different Conformations in Polymerase and Editing Mode? If a closed conformation of the exonuclease domain prohibits single strand binding, an open conformation is required for editing. The observed closed conformation may represent the enzyme in "polymerase" mode. Preliminary analysis of the crystal structure of *Tgo* pol in the low-salt conditions indicates a structural change at the interface of exonuclease and thumb, possibly reflecting a transition between open and closed forms. The closed conformation observed here may, however, be a nonphysiological artifact of the high ionic strength used for crystallization. Crystal structures of the enzyme in both polymerase and editing modes are required.

Adaptation to High Temperatures. *Tgo* is a sulfur-metabolizing, extremely thermophilic organism, with a growth range between 55°C and 95°C. For accurate replication at this temperature range, the polymerase must not only be stable, but must also adequately bind substrate DNA. A comparison with gp43 from the mesophilic bacteriophage RB69 indicates several such adaptations to high temperatures. Several loops are shorter in *Tgo* pol than in gp43 (Fig. 2), and there is an increase in hydrogen bonded β-strand content: *Tgo* pol secondary structure includes 43% helix, 22% β-strands, and 19% turns (calculated according to ref. 58), whereas gp43 has 42% helix, 17% β-strands, and 19% turns.

Although rare among cytoplasmic or nuclear proteins, two disulfide bridges might be formed: cysteine pairs 428–442 and

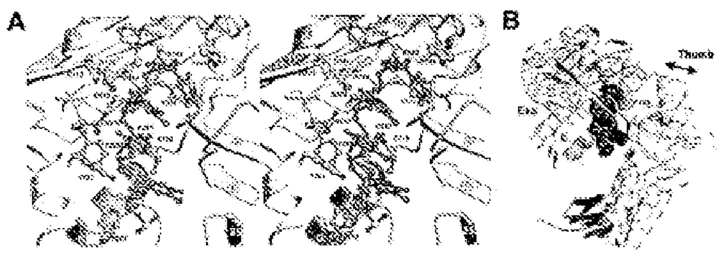


FIG. 5. 3′ → 5′ exonuclease active site. (A) Stereoribbon representation with modeled DNA using the color code of Fig. 4A (ball-and-stick) and Fig. 2 (ribbons). Active-site residues are shown as ball-and-stick representation. The single-stranded DNA has been taken from the coordinates of RB69-single-stranded DNA complex (23). The orientation of the DNA has been obtained by superimposing the DIEDQ motif in *Tgo* pol with corresponding motif of RB69 gp43 (DIE114). Strand 27 and its preceding loop from the thumb (green) is apparently in collision with the modeled DNA. (B) Comparison of the exonuclease–thumb interface between *Tgo* pol (color code of Fig. 2B) and RB69 gp43 (gray). In *Tgo* pol, the lid of the editing site (red) is bent outward compared with the equivalent loop of gp43 (yellow), allowing the tip of the thumb to move several Å (5) closer to the exonuclease. This conformation is incompatible with formation of an editing complex (the p/dTN of gp43 is shown as brown space-filling model).

506–509, although oxidized, are poised for attachment (Fig. 3). Model refinement and electron density inspection with and without constraints for the disulfide bonds verified the reduced state (observed unrestrained SG–SG distances: 2.8 Å and 3.0 Å). This is consistent with our E. coli expression and further rules out structural perturbation by oxidative oxidation. These cysteines are located in the palm domain and are conserved among B type enzymes from hyperthermophilic sulfur-metabolizing archaeons, but not among mesophilic homologs (Fig. 3). The Cys-428–Cys-442 bridge stabilizes the compact fold of the loop segment between helix N in the palm domain and helix O in the fingers and presumably also the relative orientation of these helices. In addition, the loop segment packs against helix Q in the palm domain. Helix Q, the spine of the palm domain, is further stabilized at the first helical turn by the second disulfide bridge between Cys-506 and Cys-509.

A much enhanced complementary positive potential for all three DNA-binding clefts of Tgo pol is observed relative to gp43 (Fig. 2). Thus, in addition to hydrogen bonding and specific DNA–protein interactions, binding to Tgo pol has an additional strong stabilizing electrostatic component.

An increase in the number of salt bridges is often associated with thermostability. Although Tgo pol has a greater total number of charged residues (262) than gp43 (245), both molecules have 54 salt bridges within a 3–5 Å bound. However, in the 5–7 Å range of charge distance, Tgo pol has 77 ion pairs compared with 45 for gp43. This large increase results in a more highly charged surface of Tgo pol, accompanied by a more balanced charge distribution, compared with gp43 where charges are often located in patches (Fig. 7).

We thank Dr. Uwe Jacob and Martin Augustin for stimulating discussions, Dr. Nediljko Budisa for help with expression of selenomethionine variants, and Dr. Hans Bartunik and Dr. Gleb Bourenkov at DESY, Hamburg for help with data collection.

1. Waga, S. & Stillman, B. (1998) Annu. Rev. Biochem. 67, 721–751.
2. Maga, G. & Hübscher, U. (1996) Biochemistry 35, 5764–5777.
3. Wang, T. S. F. (1991) Annu. Rev. Biochem. 60, 513–552.
4. Kornberg, A. & Baker, T. A. (1992) DNA Replication (Freeman, New York).
5. Joyce, C. M. & Steitz, T. A. (1994) Annu. Rev. Biochem. 63, 777–822.
6. Perler, F. B., Kumar, S. & Kong, H. (1996) Adv. Protein Chem. 48, 377–435.
7. Brautigam, C. A. & Steitz, T. A. (1998) Curr. Opin. Struct. Biol. 8, 54–63.
8. Arnold, E., Ding, J., Hughes, S. H. & Hostomsky, Z. (1995) Curr. Opin. Struct. Biol. 5, 27–38.
9. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985) Nature (London) 313, 762–766.
10. Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R. & Steitz, T. A. (1988) Proc. Natl. Acad. Sci. USA 85, 8924–8928.
11. Kiefer, J. R., Mao, C., Hansen, C. J., Basehore, S. L., Hogrefe, H. H., Braman, J. C. & Beese, L. S. (1997) Structure (London) 5, 95–108.
12. Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998) Nature (London) 391, 304–307.
13. Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W. & Steitz, T. A. (1995) Nature (London) 376, 612–616.
14. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998) Nature (London) 391, 251–258.
15. Davies, J. F., Almassy, R. J., Hostomska, Z., Ferre, R. A. & Hostomsky, Z. (1994) Cell 76, 1123–1133.
16. Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H. & Kraut, J. (1994) Science 264, 1930–1935.
17. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994) Science 264, 1891–1903.
18. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992) Science 256, 1783–1790.
19. Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., et al. (1993) Proc. Natl. Acad. Sci. USA 90, 6320–6324.
20. Jager, J., Smerdon, S. J., Wang, J., Boisvert, D. C. & Steitz, T. A. (1994) Structure (London) 2, 869–876.
21. Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997) Cell 89, 1087–1099.
22. Edgell, D. R. & Doolittle, W. F. (1997) Cell 89, 995–998.
23. Cann, I. K. O., Komori, K., Toh, H., Kanai, S. & Ishino, Y. (1998) Proc. Natl. Acad. Sci. USA 95, 14250–14255.
24. Debyser, Z., Tabor, S. & Richardson, C. C. (1994) Cell 77, 157–166.
25. Nossal, N. G. (1994) Molecular Biology of Bacteriophage T4, ed. Karam, J. D. (Am. Soc. Microbiol., Washington, DC), pp. 43–53.
26. Nossal, N. G. (1992) FASEB J. 6(3), 871–878.
27. Fanning, E. & Knippers, R. (1992) Annu. Rev. Biochem. 61, 55–85.
28. Kelman, Z. & O'Donnell, M. (1995) Annu. Rev. Biochem. 64, 171–200.
29. De Pamphilis, M. L., ed. (1996) DNA Replication in Eukaryotic Cells (Cold Spring Harbor Lab. Press, Plainview, NY).
30. Waga, S. & Stillman, B. (1994) Nature (London) 369, 207–212.
31. Baker, T. A. & Bell, S. P. (1998) Cell 92, 295–305.
32. Uemori, T., Sato, Y., Kato, I., Doi, H. & Ishino, Y. (1997) Genes Cells 2(8), 499–512.
33. Braithwaite, D. K. & Ito, J. (1993) Nucleic Acids Res. 21(4), 787–802.
34. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D. & Venter, J. C. (1996) Science 273, 1058–1073.
35. Miroshnichenko, M. L., Gongadze, G. M., Rainey, F. A., Kostyukova, A. S., Lysenko, A. M., Chernyh, N. A. & Bonch-Osmolovskaya, E. A. (1998) Int. J. Syst. Bacteriol. 48, 23–29.
36. Bonch-Osmolovskaya, E., Svetlichny, V., Ankenbauer, W., Schmitz-Agheguian, G., Angerer, B., Ebenbichler, C. & Laue, F. (1998) Eur. Patent Appl. EP 0 834 570 A1.
37. Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. & Huber, R. (1995) Eur. J. Biochem. 230, 788–796.
38. Leslie, A. G. W. (1996) MOSFLM User Guide (MRC Laboratory of Molecular Biology, Cambridge, UK), Version 5.5.
39. Otwinowski, Z. & Minor, W. (1997) Methods Enzymol. 276, 307–326.
40. Collaborative Computational Project, No. 4 (1994) Acta Crystallogr. D 50, 760–763.
41. Turk, D. (1992) Ph.D. Thesis (Technische Universität, München).
42. Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1997) Proc. Natl. Acad. Sci. USA 94, 5018–5023.
43. Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H. & Steitz, T. A. (1996) Biochemistry 35, 8110–8119.
44. Sattar, A. K., Lin, T. C., Jones, C. & Konigsberg, W. H. (1996) Biochemistry 35, 16621–16629.
45. Reha-Krantz, L. J. & Nonay, R. L. (1994) J. Biol. Chem. 269, 5635–5643.
46. [illegible] (1998) Proc. Natl. Acad. Sci. USA 95, [illegible].
47. Lin, T. C., Karam, G. & Konigsberg, W. H. (1994) J. Biol. Chem. 269, 19286–19294.
48. Kabsch, W. & Sander, C. (1983) Biopolymers 22, 2577–2637.

Appendix I

We have purified and characterized the Family BDNA polymerase from the archaeon Methanococcus maripaludis, cloned from ATCC 43000. This polymerase has a 41% sequence identity and 63% sequence similarity with Vent DNA Polymerase when analyzed using NCBI Blast 2 and the default parameters.

We performed the titration assay described in Example 1 of the patent application, using the Mma, Vent (exo-), and 9°N (exo+) DNA Polymerases. Experimental details and data are given in the attached figure.

For each of the three polymerases, a comparison of lanes using dideoxyCTP (ddCTP) with those using equivalent concentrations of acycloCTP (acyCTP) reveals shorter products in lanes utilizing acyCTP. These shorter products result from more efficient insertion of the acyCTP terminator compared to incorporation of the ddCTP terminator. Thus, all three polymerases incorporated acyCTP more efficiently than ddCTP.

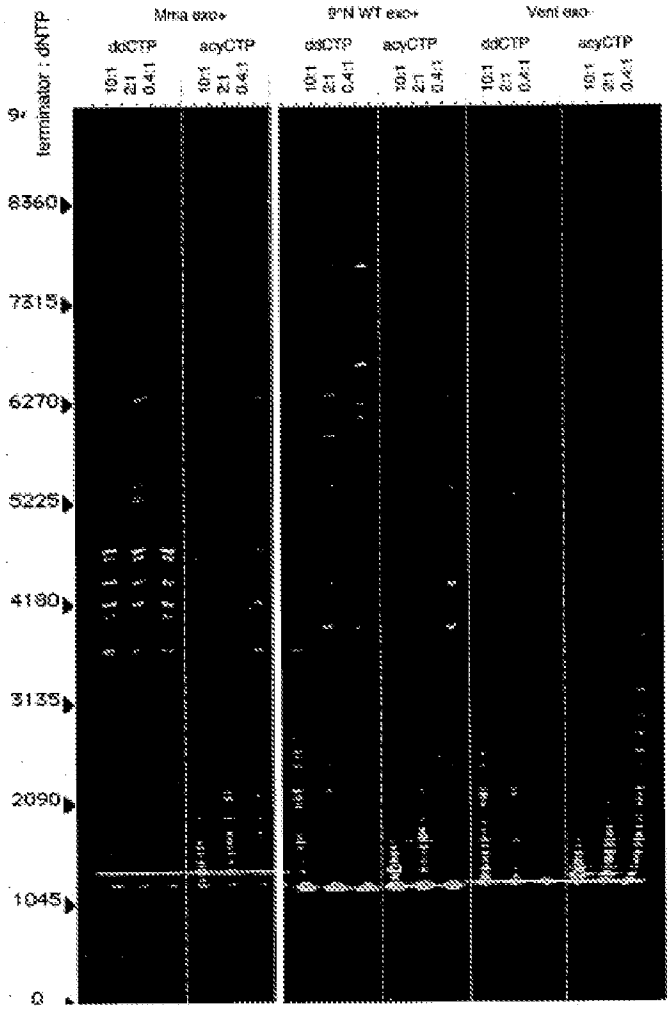
Figure Legend

The ability of acyNTPs and ddNTPs to act as chain terminators was tested using a titration assay of the type described in Example 1. Incorporation of ddCTP was compared to that of acyCTP, respectively, using Methanococcus maripaludis DNA polymerase, 9°N (exo+) DNA polymerase and Vent® (exo-) DNA polymerases.

Incorporation of ddCTP and acyCTP was assayed by mixing 8 µl of reaction cocktail (0.025 µM 5' [FAM] end-labeled #1224-primed M13mp18, 62.5 mM NaCl, 12.5 mM Tris-HCl (pH 7.9 at 25°C), 12.5 mM $MgCl_2$, 1.25 mM

dithiothreitol, *Methanococcus maripaludis* DNA polymerase or 0.125 U/µl 9°N (exo+) DNA polymerase or 0.125 U/µl Vent® (exo-) DNA polymerase) with 2 µl of 5X nucleotide analog/nucleotide solution to yield the final ratios of analog:dNTP indicated in the figures. After incubating at 72°C for 20 minutes, the reactions were halted by the addition of 10 µl formamide. Samples were then heated at 72°C for 3 minutes and a 1 µl aliquot was loaded on a 4% polyacrylamide urea gel and detected by an ABI377 automated DNA sequencer.

ddCTP v. acyCTP incorporation by archaeal DNAPs



BEST AVAILABLE COPY